A Study on Numerical Methods for Scalar Conservation Laws

Dissertation submitted to
Visvesvaraya National Institute of Technology, Nagpur
in partial fulfillment of the requirements for the award of the degree

Master of Science

in

Mathematics

by

Sagar Bharat Berad

(MS23MTH007)

under the guidance of

Dr. G. Naga Raju

Department of Mathematics
Visvesvaraya National Institute of Technology
Nagpur - 440010, (India)

May, 2025

A Study on Numerical Methods for Scalar Conservation Laws

Dissertation submitted to
Visvesvaraya National Institute of Technology, Nagpur
in partial fulfillment of the requirements for the award of the degree

Master of Science
in
Mathematics

by

Sagar Bharat Berad

(MS23MTH007)

under the guidance of

Dr. G. Naga Raju

Department of Mathematics
Visvesvaraya National Institute of Technology
Nagpur - 440010, (India)

May, 2025

© Visvesvaraya National Institute of Technology (VNIT) 2009

Department of Mathematics

Visvesvaraya National Institute of Technology, Nagpur

Declaration

I, Sagar Bharat Berad, hereby declare that this dissertation titled "A Study on Numerical Methods for Scalar Conservation Laws" is carried out by me in the Department of Mathematics Visvesvaraya National Institute of Technology, Nagpur. The work is original and has not been submitted earlier whole or in part for the award of any degree/diploma at this or any other Institution/University.

Name and Signature
Date:

Certificate

This is to certify that the dissertation titled "A Study on Numerical Methods for Scalar Conservation Laws" is submitted by **Sagar Bharat Berad** in partial fulfillment of the requirements for the award of the degree of **Master of Science in Mathematics**, VNIT Nagpur. The work is comprehensive, complete and fit for final evaluation.

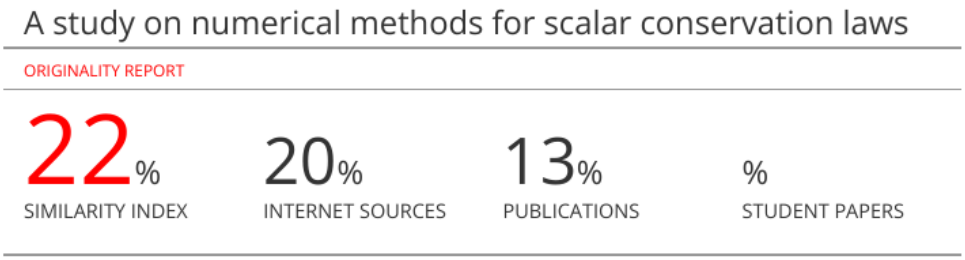
Dr. G. Naga Raju
Associate Professor, Mathematics,
VNIT, Nagpur

Head, Department of Mathematics
VNIT, Nagpur
Date:

VISVESVARAYA NATIONAL INSTITUTE OF TECHNOLOGY NAGPUR

Non-Plagiarism Certificate

Certified that M.Sc. Dissertation titled "**A Study on Numerical Methods for Scalar Conservation Laws**" submitted by **Sagar Bharat Berad**, Enrollment No. **MS23MTH007**, ID No. **30127**, is original in nature and not copied from any literature material. Due to credit has been given to the original researcher wherever referred.



Signature
Name of the Student: **Sagar Bharat Berad**
Enrollment No: MS23MTH007

Dr. G. Naga Raju
Associate Professor
Department of Mathematics,
VNIT, Nagpur

Acknowledgements

First and foremost, I would like to express my profound gratitude to my supervisor, **Dr. G. Naga Raju**, for their invaluable guidance, insightful feedback, and unwavering support throughout the duration of this research. Their expertise and encouragement have been instrumental in shaping the direction and quality of this work. I would also like to extend my sincere thanks to the faculty members and staff of the **Department of Mathematics**, **VNIT Nagpur**, for their support and for providing an academic environment conducive to research and learning. My appreciation is further extended to my colleagues and peers, whose collaboration, discussions, and moral support have significantly contributed to the progress of this thesis. Lastly, I am deeply thankful to my family for their constant encouragement, patience, and understanding throughout the course of my academic journey. Their unwavering belief in me has been a source of strength and motivation.

Abstract

This study presents numerical methods for solving conservation laws, which are key to modeling real-world phenomena like fluid flow, traffic, and heat transfer. Since exact solutions are rarely available especially in the presence of shocks or some other nonlinearities efficeient numerical schemes are essential. The study begins with linear transport equations and extends to nonlinear scalar conservation laws. It explores finite difference and finite volume methods, including schemes like Lax-Friedrichs, Lax-Wendroff, Beam-Warming, and Godunov. Emphasis is placed on stability, convergence, and handling discontinuities using weak solutions and entropy conditions. Numerical experiments demonstrate the effectiveness of the methods in capturing complex solution behavior accurately.

Contents

1 Introduction **2**
1.1 Characteristic method 4

2 Finite difference schemes **12**
2.1 Central Difference Method 13
2.2 Lax–Friedrichs Method 15
2.3 Lax–Wendroff Method 17
2.4 The Beam–Warming Method 18

3 Finite volume schemes **19**
3.1 Finite volume method 19
3.2 Consistency, Conservation and Monotonicity 28

4 Second-Order Schemes **34**
4.1 The REA algorithm 38
4.2 The Minmod Limiter 42
4.3 Another limiters 44
4.4 Temporal Discretization 46

5 Conclusion **48**

Chapter 1

Introduction

Conservation laws are a class of partial differential equations (PDEs) that express the conservation of physical quantities like mass, momentum, or energy. These equations are fundamental in modeling real-world phenomena in fluid dynamics, traffic flow, gas dynamics, weather systems, and more.

Analytical solutions to conservation laws are often not available, especially when shocks (discontinuities) or complex geometries arise. Numerical methods allow us to approximate solutions to these equations on computers by discretizing time and space.

Suppose $\Omega \subset \mathbb{R}^n$ is the domain of interest, and u represents a physical quantity defined over this domain. This quantity could signify temperature, pressure, population density, or similar metrics. A fundamental observation can be made:

The rate at which u changes within a subregion $\omega \subset \Omega$ equals the net generation or loss within ω plus the flux of u across the boundary $\partial\omega$.

This is mathematically described as

$$\frac{d}{dt}\int_{\omega} \mathbf{u}(x,t)\,dx = \int_{\omega} \mathbf{S}(x,t)\,dx - \int_{\partial\omega} \mathbf{F}(u)\cdot\nu\,d\sigma(x)$$

Here ν is outward unit normal, F denotes the flux, and S represents sources or sinks. Applying the divergence theorem simplifies this to

$$\frac{d}{dt}\int_{\omega} \mathbf{u}\,dx = \int_{\omega} (\mathbf{S}(x,t) - \nabla\cdot\mathbf{F}(u))\,dx$$

Since the above equation hold for every sub-domain ω of Ω, we may proceed with an infinitesimal to get the following differential equation

$$\mathbf{u}_t + \nabla\cdot\mathbf{F} = \mathbf{S} \qquad \forall\ (x,t)\ \in\ (\Omega, \mathbb{R}_+)$$

This differential equation is known as a balance law.

Examples

Scalar diffusion equation

Consider $\mathbf{u} = u(\mathbf{x}, t)$ as the temperature at a position $\mathbf{x}$ and moment t. Fourier's law describe heat transfer by stating that the flow of heat moves in the direction opposite to the temperature gradient and it,s magnitude and proportional to that gradient, i.e.

$$\mathbf{F} = -k\nabla u,$$

where $k > 0$ represents the thermal conductivity characteristic of the material. Since there is no internal heat generation or absorption, the source term is equal to zero.Therefore the conservation law becomes

$$u_t - \operatorname{div}(k\nabla u) = 0.$$

If k is the same conductance and the problem is one-dimensional, then the equation reduces to

$$u_t - ku_{xx} = 0.$$

This is the heat equation, describing temperature diffusion in a uniform medium.

Advection equation

Let $u = u(x, t)$ represent the *density of cars* on a road. Assume that cars move with a velocity $a(x, t)$, The traffic flow of cars is then given by

$$\mathbf{F} = a(x, t)u.$$

Assuming that there are no cars entering or leaving the road (i.e., no sources or sinks), the conservation law become

$$u_t + \operatorname{div}(a(x, t)u) = 0.$$

In the one-dimensional case and for constant velocity a, the above equation simplifies to

$$u_t + au_x = 0. \tag{1.1}$$

This is the *advection* equation, modeling the movement of traffic density along the road.

Burgers' equation

The flux for the Burger's equation is given by

$$\mathbf{F} = \frac{u^2}{2}.$$

Substitute this into general conservation law form

$$u_t + \operatorname{div}\left(\frac{u^2}{2}\right) = 0.$$

In one spatial dimension, the divergence operator simplifies to a spatial derivative with respect to x resulting in

$$u_t + \left(\frac{u^2}{2}\right)_x = 0. \tag{1.2}$$

This is inviscid Burgers' equation. More generally, scalar conservation laws take the form

$$u_t + f(u)_x = 0 \tag{1.3}$$

where $f(u)$ is a nonlinear flux function. [**2**]

In what follows we study the solution behavior of scalar Advection and Burger's equations.

1.1 Characteristic method

Consider the advection equation (1.1) with the initial data, [**1**]

$$u(x,0) = u_0(x) \quad \forall x \in \mathbf{R}.$$

The idea is that we have to reduce a PDE into an ODE. along curves where the solution remains constant.

$$\begin{aligned}\frac{d}{dt}u(x(t),t) &= \frac{\partial u}{\partial t}(x(t),t) + \frac{d}{dt}x(t).\frac{\partial u}{\partial x}(x(t),t)\\ &= u_t + a(x,t)u_x\\ &= 0\end{aligned}$$

Since we assumed u is the solution of advection equation.
Implies that , if $x(t)$ meets the ODE

$$\frac{d}{dt}x(t) = a(x,t) \tag{1.4}$$

$$x(0) = x_0$$

The solution $x(t)$ of this equation is called *characteristics curve.*
Now if we take the velocity field $a(x,t) \equiv a$, where a is constant, an explicit solution to the above mentioned characteristic equation is obtained as

$$x(t) = x_0 \ + \ at$$

The characteristic equation is $x_0 = x - at$. Hence the solution of (1.1) is become

$$u(x,t) = u_0(x-at); \quad \forall (x,t) \ \in \mathbb{R} \times \mathbb{R}_+$$

The resultant formula suggests that the speed a is used to convey the starting information.

Lemma 1. *If $u(x,t)$ is a continuous solution of $u_t + a(x,t)u_x = 0$, then $a \in C^1\ (\mathbb{R}, \mathbb{R}_+)$ breaks down to zero at infinite, that is, $\lim_{|x|\to\infty} u(x,t) = 0$ for all $x \in \mathbb{R}, t > 0$. Thus, u fulfills this energy bound.*

$$\int_{\mathbb{R}} u^2(x,t)dx \leq e^{||a||_{C^1}t} \int_{\mathbb{R}} u_0^2(x)dx \ ; \quad \forall\ t > 0$$

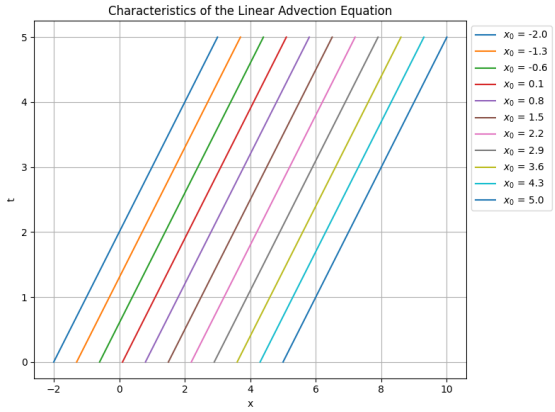


Figure 1.1: Characteristic curve $x(t)$ for advection equation (1.1)

Proof. By multiplying by u on each side the equation, we get

$$uu_t + a(x,t)uu_x = 0$$

$$\left(\frac{u^2}{2}\right)t + a(x,t)\left(\frac{u^2}{2}\right)x = 0$$

$$\left(\frac{u^2}{2}\right)t + \left(a(x,t)\frac{u^2}{2}\right)x = a_x(x,t)\left(\frac{u^2}{2}\right)$$

$$\int_{\mathbb{R}} \frac{d}{dt}\left(\frac{u^2}{2}\right)dx + \int_{\mathbb{R}} \frac{d}{dx}\left(a(x,t)\frac{u^2}{2}\right)dx = \int_{\mathbb{R}} a_x(x,t)\left(\frac{u^2}{2}\right)dx$$

$$\frac{d}{dt}\int_{\mathbb{R}}\left(\frac{u^2}{2}\right)dx + \left(a(x,t)\frac{u^2}{2}\right)_{-\infty}^{\infty} = \int_{\mathbb{R}} a_x(x,t)\left(\frac{u^2}{2}\right)dx$$

Using Gronwall's Inequality -

$$\int_{\mathbb{R}} \frac{u^2(x,t)}{2}dx \leq e^{||a||_{C^1}t}\int_{\mathbb{R}} \frac{u_0^2(x)}{2}dx$$

This energy estimate shows that energy of the solution to the Linear advection equation (1.1) are Bounded. □

Trajectories determined by Burgers' equation

To solve Burgers' equation, we again consider the method of characteristics but here the speed is depend upon solution itself. For smooth initial conditions $u_0(x)$, the characteristics equation is refer (1.4)

$$\frac{dx}{dt} = u(x,t)$$

$$x(0) = x_0$$

$$x(t) = u_0(x_0)t + x_0.$$

We start by considering the initial data

$$u_0(x) = \begin{cases} u_L & \text{if } x < 0 \\ u_R & \text{if } x > 0 \end{cases} \tag{1.5}$$

This type of data is rather straightforward and is made up of constants that are separated at the origin by a discontinuity. This setup is known as the Riemann data. The solution u is constant along characteristics by definition, meaning that, since $u(x,t) = u_0(0)$, the solution to equation (1.5) remains constant in regions where u_0 is uniform.

In particular take $u_L = 1$ and $u_R = 0$ in (1.5). If $x_0 < 0$ then the speed is $u_0(x_0) = 1$ and if $x_0 > 0$ then the speed is 0. The characteristics diverge or intersect depending on the initial values. In this example, characteristics cross, indicating a breakdown of classical solutions see Figure (1.2).
As the characteristic starts to merge, an issue of what happens to the solution now occurs. And how can one define a solution in this situation?

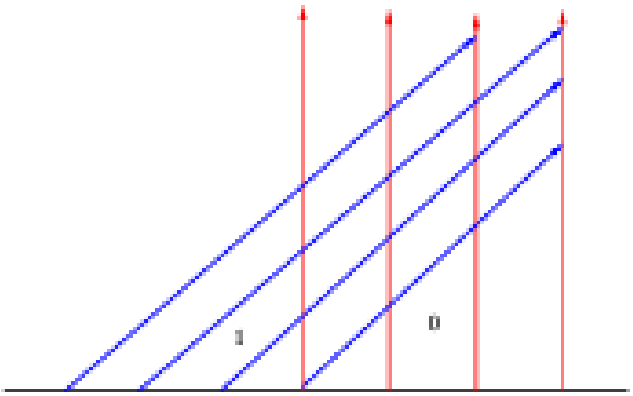


Figure 1.2: Characteristics intersect for $(u_L, u_R) = (1, 0)$

Look at the original data,

$$u(x, 0) = \sin(\pi x)$$

within the range $[-1, 1]$. The strange behavior of characteristics shows that even with smooth initial data, smooth solutions for the conservation equation (1.3) cannot be found. The solution moves with the velocity $u_0(x)$ at each point x, according to a heuristic interpretation of the characteristic equation. As a result, the characteristics method suggests that the solution acts as Figure (1.3) illustrates. The wave expands in one part and compresses in another. In particular, the solution can be multi-valued. This further suggests that smooth solution of (1.3) do not exist.

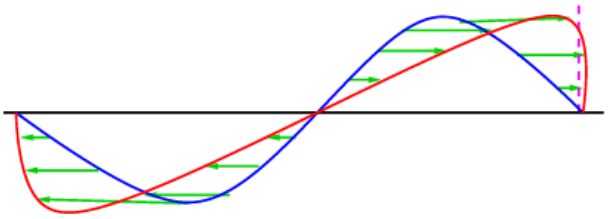


Figure 1.3: A multivalued solution

Weak Solutions

Since classical (smooth) solutions break down, we introduce weak solutions, which satisfy the conservation law in an integrated sense. Assume for the moment that the *conservation law* has a smooth solution. Let $\phi \in C_c^1(\mathbb{R} \times \mathbb{R}_+)$ represent the set of all function from A to $\mathbb{R}$ that are continuously differentiable with compacted support, and ϕ be a smooth function. When we integrate the conservation law (1.3) by parts over the domains $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$, we obtain

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} (u\phi_t + f(u)\phi_x)\, dtdx \; + \int_{\mathbb{R}} u_0(x)\phi(x,0)\, dx = 0 \tag{1.6}$$

A function $u \in L^\infty$ that satisfies this identity for all test functions is a weak solution.

Definition 1.1. *(Weak solution) A function with the value $u \in L^\infty(\mathbb{R} \times \mathbb{R}_+)$ provided that the identity* (1.6) *is satisfied for all test function $\phi \in C_c^1(\mathbb{R} \times \mathbb{R}_+)$,then is a weak solution of* (1.3).

The Rankine-Hugoniot condition

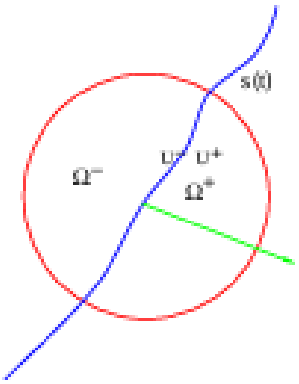


Figure 1.4: across the shock

Let the shock wave be defined by the curve $x = \gamma(t)$. Let assume that $u \in C^1(\Omega^+)$ and $u \in C^1(\Omega^-)$. Let $u^+(t)$ and $u^-(t)$ represent the values on either side of the shock.

Integrating (1.6) by parts over the Ω , we get

$$\int_{\Omega}(u\phi_t + f(u)\phi_x)d\Omega = 0$$

let (u,f) = ω and $(\phi_t, \phi_x) = \nabla\phi$

$$\int_{\Omega}(w\nabla\phi)d\Omega = 0$$

$$\int_{\Omega^-}(w\nabla\phi)d\Omega + \int_{\Omega^+}(w\nabla\phi)d\Omega = 0$$

using Green's Theorem the above equation becomes,

$$\int_{\Omega^-}(wn\phi ds) - \int_{\Omega^-}(\Delta w\phi d\Omega^-) + \int_{\Omega^+}(wn\phi ds) - \int_{\Omega^+}(\Delta w\phi d\Omega^+) = 0$$

Where n is the outward unit normal.

$$\int_{\Omega^-}(u^-, f(u^-))(s(t), -1)\phi ds + \int_{\Omega^+}(u^+, f(u^+))(-s(t), 1)\phi ds - \int_{\Omega^+\cup\Omega^-}(u_t + f(u)_x)\phi d\Omega = 0$$

Now we have,

$$\int_{\partial\Omega}[(u^- - u^+)s(t) + (f(u^+) - f(u^-))]\phi ds = 0$$

Because ϕ is an arbitrary test function, the integrand of the remaining integral must be identically zero . It follow that the shock velocity must obey

$$\gamma'(t) = \ s(t) \ = \frac{f(u^+) - f(u^-)}{(u^+ - u^-)}$$

This state is known as the *Rankine-Hugoniot* state.

Riemann problem

Returning to the example $u_L = 1$ and $u_R = 0$. Due to the intersection of characteristics, a continuous solution cannot be formed . The weak solution consists of a shock propagating at speed determined by the Rankine-Hugoniot condition.

$$s(t) = \frac{\frac{u_R^2}{2} - \frac{u_L^2}{2}}{u_R - u_L} = \frac{1}{2} \tag{1.7}$$

As a result, the weak solution looks like

$$u(x,t) = \begin{cases} 1 & \text{if } x < \frac{1}{2}t \\ 0 & \text{if } x > \frac{1}{2}t \end{cases} \tag{1.8}$$

It is easy to confirm that (1.8) fulfills *Rankine-Hugoniot* since the solution's geometry shows that the characteristics flow within the shock (see Figure(1.5)). Each point in the

plane is thus covered by a characteristic, and we can relate a characteristic to the original data for each of them, indicating that the starting data dictates the complete solution.

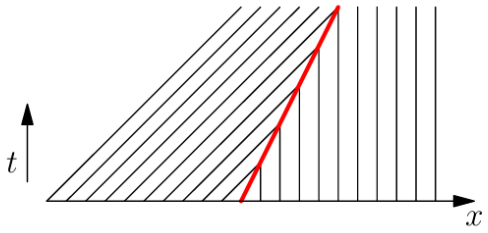


Figure 1.5: Riemann problem characteristic (1.8)

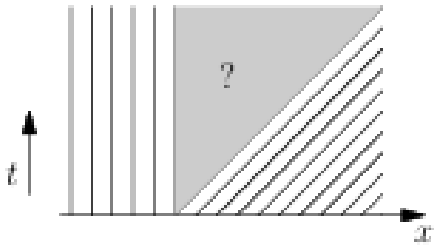


Figure 1.6: Riemann problem characteristic (1.9)

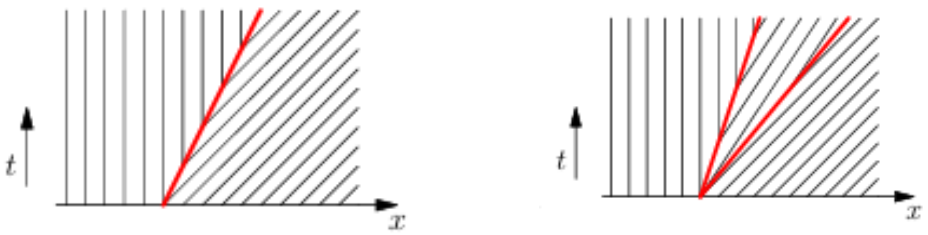
Alternatively, if $u_L = 0$ and $u_R = 1$. As in the preceding problem, if we keep to the characteristics obtained from the x-axis, we now obtain an area devoid of characteristics; refer to Figure (1.7).There are a number of approaches to fill in the gaps in this region. We can conclude that one probable weak answer is offered by using the Rankine-Hugoniot condition.

$$u(x,t) = \begin{cases} 0 & \text{if } x < \frac{1}{2}t \\ 1 & \text{if } x > \frac{1}{2}t \end{cases} \tag{1.9}$$

As seen in image (1.7)(a), this solution has only one red-colored curve. But this isn't the only available answer, infinitely many weak solutions exist, depending on how the intermediate states are chosen. Let's take u_m and applying the Rankine-Hugoniot condition yields the weak solution

$$u(x,t) = \begin{cases} 0 & \text{if } x < \frac{u_m}{2}t \\ u_m & \text{if } \dfrac{u_m}{2}t < x < \dfrac{1+u_m}{2}t \\ 1 & \text{if } x > \frac{1+u_m}{2}t \end{cases} \tag{1.10}$$

The characteristics are shown in (1.7)(b). Similarly, the *Rankine-Hugoniot* condition can be used to create an endless number of weak solutions with various intermediate states. The definition of weak solutions implicitly addresses this issue of non-uniqueness. Since these solutions aren't always original, certain additional requirements must be put in place.To find these additional criteria, We observe that the characteristics of both flows emanate from the shock; refer to Figure (1.7).



(a) answer (1.9)

(b) answer (1.10)

Figure 1.7: Characteristics of several Riemann problem weak solutions (1.9), red highlights indicate discontinuity.

On the other hand, as shown in Figure (1.5), the characteristics of the solution (1.8) transfer into the shock. Characteristics are a representation of the information flow. The initial data should always provide the information for an evolution equation. This is clearly seen by the weak answer (1.8). Nevertheless, for the weak solution given by (1.9) and (1.10), it appears that information originates at the shock. Conditions on the features across a shock can be used to define this intuitive condition ,stating that information originates form the original data and is not created during the shock. Consider the states on both side of a shock moving at speed $s(t)$ are $u^-(t)$, $u^+(t)$. The condition can enforce the requirement that Burger's equation characteristics flow into the shock and that information is taken from the beginning line.

$$u^- > s(t) > u^+$$

For general convex flux f, the Lax entropy condition requires that:

$$f'(u^-) > s(t) > f'(u^+).$$

This ensures information flows into the shock, not out. *Lax entropy condition* is the term for this.

Rarefaction waves

For the remainder of this section, we consider the case where the flux function f is strictly convex. A continuous solution to equation (1.3) is constructed under this assumption. We observe that scaling the variable x and t by a factor $\lambda/$ i.e. replacing them with λx and λt, ;eaves the equation invariant- meaning a solution remains valid under this transformation. Since the initial Riemann data (1.5) is also invariant under the scaling $x \Rightarrow \lambda x$, it is natural to consider a self-similar solution, implies that the solution of Riemann problem is depends only on the ration $\frac{x}{t}$. Let,

$$u(x,t) = v(x/t)$$

put, $\eta = x/t$, where η is the symmetry variable.

$$
\begin{aligned}
&v(\eta)_t + f(v(\eta))_x = 0 \\
&v_\eta\ \eta_t + f'(v(\eta))v_\eta\ \eta_x = 0 \\
&v_\eta(-\frac{x}{t^2} + f'(v(\eta))\frac{1}{t}) = 0 \\
&v_\eta(f'(v(\eta)) - \eta) = 0
\end{aligned}
$$

In the non-trivial case $v_\eta \neq 0$
So we get,

$$f'(v(\eta)) = \eta$$

As a result,

$$v\left(\frac{x}{t}\right) = (f')^{-1}\left(\frac{x}{t}\right)$$

Such a self-similar solution is known as a rarefaction wave.
Let us take the Riemann problem (1.3), (1.5). When $u_L < u_R$,here the appropriate answer is a rarefaction wave instead of a shock, a continuous fan-shaped structure described by:

$$u(x,t) = \begin{cases} u_L & \text{if } x \leq f'(u_L)t \\ v\left(\frac{x}{t}\right) & \text{if } f'(u_L)t < x \leq f'(u_R)t \\ u_R & \text{if } x > f'(u_R)t \end{cases} \tag{1.11}$$

Clearly the Lax entropy condition is satisfies by the weak solution (1.11). In particular if we take Burger's equation with the Riemann data $u_L = 0$ and $u_R = 1$, the Figure (1.8) represents the solution (1.11). One can note that the rarefaction wave is aligned with the characteristic curves and it is contradict to Figure (1.7).

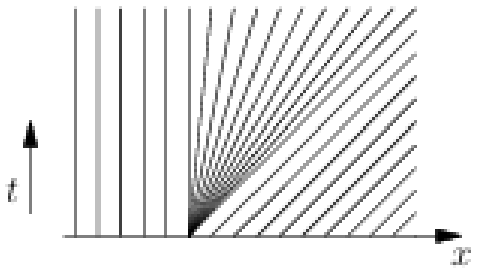


Figure 1.8: The solution representing rarefaction

We proceed by possess a formula for creating weak solution to the Riemann problem defined by equation (1.3) and (1.5) for a conservation law with a strictly convex flux function is considered. f. The answer is based upon whether $u_L < u_R$ or $u_L > u_R$. If $u_L > u_R$ then the weak solution consist two state separated by a shock satisfies entropy condition. $u_L < u_R$ then the weak solution (1.11) consist of a rarefaction wave between the two states. In any of this cases, the wave speed is bounded in absolute value by the maximum of $|f'(u_L)|$ and $|f'(u_R)|$.

Chapter 2

Finite difference schemes

When the field velocity is $a(x,t) = a$ is complex or nonlinear, it can be challenging or perhaps impossible to derive a clear method to the characteristic equation's answer. As a result, we must approximate the answer using numerical techniques. To keep things simple, we solve equation (1.1). It is simple to modify these techniques to deal with more diverse velocity fields.

Domain Discretization

Discretization of the temporal and spatial domains is the initial stage of any numerical approach. We must truncate the spatial domain $\mathbb{R}$ to a finite domain $[x_L, x_R]$ because it is unbounded. As will be covered in more detail later, this truncation necessitates the introduction of appropriate boundary constraints.

For clarity, the interval $[x_L, x_R]$ is evenly divided with a grid size of Δx. $x_0 = x_L, x_N = x_R$, and $x_{j+1} - x_j = \Delta x$ for all j are the grid points x_j that are used for this. It is simple to consider an irregular discretization of data.

We select a terminal time T for the temporal separation and split the time interval $[0, T]$ into M points. For $n = 0, 1, 2, ...M.$, this yields discrete time levels $t^n = n\Delta t$. The time and space mesh is displayed in Figure (2.1)

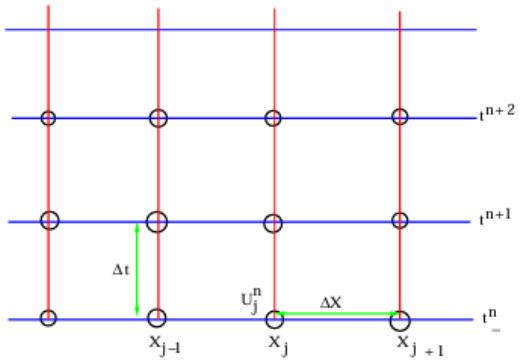


Figure 2.1: A time and space description for the mesh

2.1 Central Difference Method

We must approximate the advection equation (1.1) on the mesh. We accomplish this by substituting finite differences for the spatial and temporal derivatives. Forward differences are used in place of the time derivative, while spatial derivatives is replaced with a central difference. The resulting scheme

$$\frac{(u_j^{n+1} - u_j^n)}{\Delta t} + a\frac{(u_{j+1}^n - u_{j-1}^n)}{2\Delta x} = 0 \quad j = 1, 2, ..., N-1.$$

is known as central difference scheme. We test it on the following numerical scheme.

A numerical experiment

Using beginning data, let's consider the advection equation (1.1) in the region [0,1]

$$u_0(x) = \sin(2\pi x)$$

Assuming periodic boundary conditions makes sense because the data is periodic. So, let

$$u_0^n = u_{N-1}^n, \quad u_N^n = u_1^n.$$

exact solution computes the precise answer as $\sin(2\pi(x-at))$. With a $= 1$ and the central scheme implemented using 500 grid points, the resulting solution at $t = 2$ is illustrated in Figure (2.2). Despite being a uniform approximation, the graphical representation shows that the approach is unstable with very high oscillations.

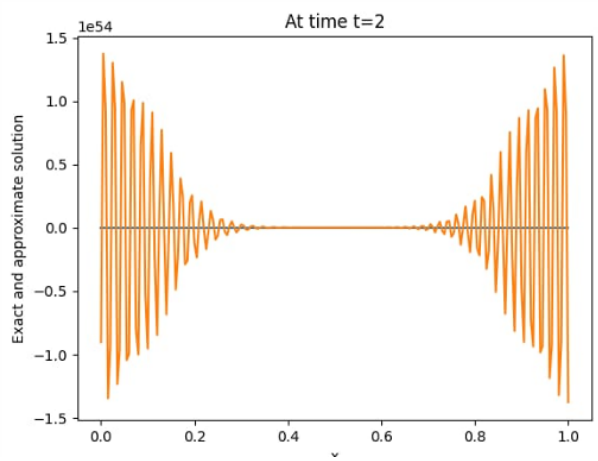


Figure 2.2: Numerical solution of (1.1) at $t = 2$ computed via the central scheme.

Explanation

Why do the results of calculations using the central scheme explode? The central scheme does appear to be a decent approximation of the transport equation, after all. The following reasoning leads to a physical explanation: Although the central difference method looks theoretically correct, it uses values from both sides of a point regardless of the direction of information flow, leading to instability.

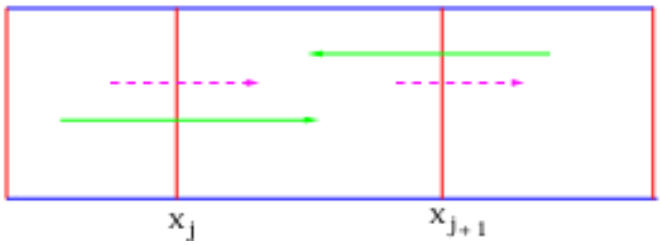


Figure 2.3: The central scheme

Von-Neumann Stability Analysis

The stability of numerical systems for solving partial differential equations (PDEs), especially linear ones, can be evaluated using Von- Neumann stability analysis.
Von-Neumann Stability Condition:

For a finite difference scheme applied to a linear PDE, the scheme is **stable** if:

$$|\xi| \leq 1 \quad \text{for all wave numbers } \beta$$

where:

- ξ is the **amplification factor**,
- β is the **wavenumber**.

Explanation:

- The solution is assumed to be of the form:

$$u_j^n = \xi^n e^{i\beta j\Delta x}$$

- Substituting into the difference scheme gives an expression for ξ.
- If $|\xi| > 1$ for any β, the error grows exponentially — the scheme is **unstable**.
- If $|\xi| \leq 1$ for all β, then the scheme is **stable** in the Von Neumann sense.

Let us consider the linear transport equation (1.1) where a is constant wave speed. We can rewrite the central scheme as

$$u_i^{n+1} = u_i^n - a\Delta t\left(\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x}\right)$$

Let us take $a > 0$ and define $\frac{a\Delta t}{\Delta x} = r$, we get:

$$u_i^{n+1} = u_i^n - \frac{r}{2}\left(u_{i+1}^n - u_{i-1}^n\right)$$

Now apply Von-Neumann stability condition we get,

$$\xi^{q+1}e^{i\beta p\Delta x} = \xi^q e^{i\beta p\Delta x} - \frac{r}{2}\left(\xi^q e^{i\beta(p+1)\Delta x} - \xi^q e^{i\beta(p-1)\Delta x}\right)$$

Canceling ξ^q and $e^{i\beta p\Delta x}$, we get:

$$\xi = 1 - \frac{r}{2}\left(e^{i\beta\Delta x} - e^{-i\beta\Delta x}\right)$$

$$\xi = 1 - \frac{r}{2}\left(2i\sin(\beta\Delta x)\right)$$

$$\xi = 1 - ir\sin(\beta\Delta x)$$

Take the complex square modulus:

$$|\xi|^2 = |1 - ir\sin(\beta\Delta x)|^2 = 1 + r^2\sin^2(\beta\Delta x)$$

Maximizing $\sin^2(\beta\Delta x) \leq 1$, we get:

$$|\xi|^2 \leq 1 + r^2$$

$$|\xi| = \sqrt{1 + r^2} > 1$$

$$\Rightarrow |\xi| > 1$$

The scheme is unstable because the amplification factor $|\xi| > 1$, indicating exponential growth in error.

Now let's move toward the another finite difference schemes which are **Forward time backward space** and **Forward time forward space** and which is given by,

- **Forward Time Backward Space (FTBS)**

$$\frac{(u_j^{n+1} - u_j^n)}{\Delta t} + a\frac{(u_j^n - u_{j-1}^n)}{\Delta x} = 0$$

- **Forward Time Forward Space (FTFS)**

$$\frac{(u_j^{n+1} - u_j^n)}{\Delta t} + a\frac{(u_{j+1}^n - u_j^n)}{\Delta x} = 0$$

With the help of Von-Neumann Stability Analysis condition it is easy to show that **Forward Time Backward Space** is **conditionally stable** and **Forward Time Forward Space** is **Unconditionally Unstable**.

2.2 Lax–Friedrichs Method

The classical Lax–Friedrichs (LxF) method has the form

$$u_j^{n+1} = \frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right) - a\Delta t\left(\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right)$$

This is quite similar to the unstable technique (Central scheme) but it replace the term u_j^n by the average $\frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right)$. The scheme remains stable for a linear hyperbolic equation as long as $|a|\frac{\Delta t}{\Delta x} \leq 1$.

Let us take same numerical example as above with time $t = 2$, $a = 1$ and the mesh points are $[100, 200, 300, 400]$ the graphical representation is given below,

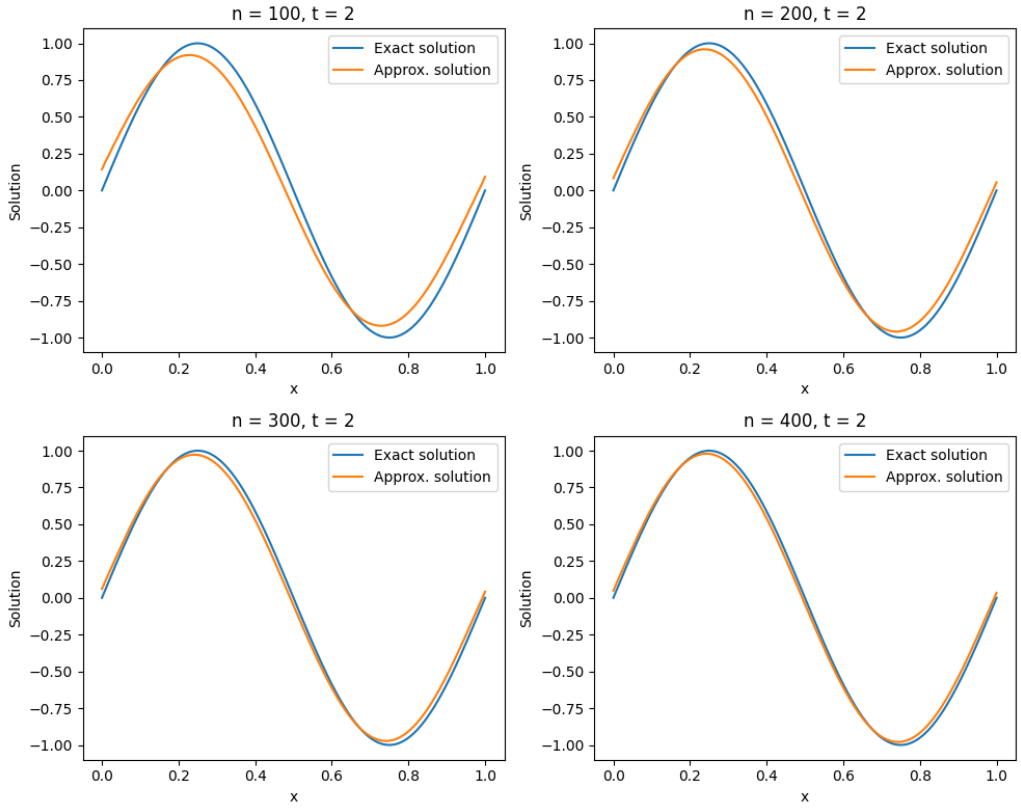


Figure 2.4: The Lax–Friedrichs scheme

Figure 2.4 shows that the Lax-Friedrichs scheme approximate the solution accurately as we increase the number of mesh points.

Using Von Neumann stability analysis, we demonstrate that the Lax-Friedrichs scheme is stable.

$$u_j^{n+1} = \frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right) - a\Delta t\left(\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right)$$

$$u_j^{n+1} = u_j^n - a\Delta t\left(\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right) + \frac{1}{2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right)$$

Substitute $\frac{a\Delta t}{\Delta x} = r$; and on simplifying the equation we've

$$u_j^{n+1} = \frac{1}{2}u_{j+1}^n(1-r) + \frac{1}{2}u_{j-1}^n(1+r)$$

Now use the Von-Neumann stability analysis and simplify we get,

$$\xi = \frac{1}{2}e^{i\beta\Delta x}(1-r) + \frac{1}{2}e^{-i\beta\Delta x}(1+r)$$

$$\xi = \frac{1}{2}\left(\cos(\beta\Delta x) + i\sin(\beta\Delta x)\right)(1-r) + \frac{1}{2}\left(\cos(\beta\Delta x) - i\sin(\beta\Delta x)\right)(1+r)$$

$$\xi = \frac{1}{2}\left(2\cos(\beta\Delta x) - 2ri\sin(\beta\Delta x)\right)$$

$$\xi = \cos(\beta\Delta x) - ri\sin(\beta\Delta x)$$

$$|\xi|^2 = \cos^2(\beta\Delta x) + r^2\sin^2(\beta\Delta x)$$

$$|\xi|^2 = 1 + (r^2-1)\sin^2(\beta\Delta x)$$

Maximizing $\sin^2(\beta\Delta x)$, we've $\sin^2(\beta\Delta x) \leq 1$

$$\Rightarrow |\xi|^2 = r$$

$$\Rightarrow |\xi| \leq 1 \quad \text{if} \quad 0 \leq \left(\frac{a\Delta t}{\Delta x}\right) \leq 1$$

Here we get $|\xi| \leq 1$, implies that the **Lax-Friedrichs scheme** is conditionally stable.

Now, the natural question arise that what is $\left(\frac{a\Delta t}{\Delta x}\right)$?

CFL Condition: Convergence of numerical method is possible only when, even in the limit as t and x approach zero, its numerical domain of reliance includes the PDE's genuine domain of dependency..

It is important to realize that the CFL condition is just required for stability.

$$\nu = |a|\left(\frac{\Delta t}{\Delta x}\right) \leq 1$$

where, The ratio ν is sometimes called CFL number or more frequently the Courant number.

Consistency+Stability $\Leftrightarrow$ Convergence.

2.3 Lax–Wendroff Method

This method introduces averaging to stabilize the scheme. It modifies the update formula and, under the CFL condition, remains stable. It's effective for reducing oscillations compared to the central scheme. The Lax–Wendroff method for the linear system is based on the Taylor series expansion.

For a grid point u_j^n (value at position j and time step n), the Lax-Wendroff update formula is:

$$u_j^{n+1} = u_j^n - \frac{a\Delta t}{2\Delta x}(u_{j+1}^n - u_{j-1}^n) + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n).$$

Where,

- Δt = time step
- Δx = spatial step
- u_j^n = value at space index j, time n

This is known as the Lax–Wendroff Method, this method is Second-order accurate in both space and time and is stable with CFL condition $\frac{a\Delta t}{\Delta x} \leq 1$.

2.4 The Beam–Warming Method

An implicit, second-order method that performs better in cases involving sharp gradients or discontinuities. It belongs to the predictor-corrector class of schemes. This is an **implicit** method and uses **backward differences** in space. The update formula for the (1.1) equation is:

$$u_j^{n+1} = u_j^n - \frac{a\Delta t}{2\Delta x}\left(3u_j^n - 4u_{j-1}^n + u_{j-2}^n\right) + \frac{a^2\Delta t^2}{2\Delta x^2}\left(u_j^n - 2u_{j-1}^n + u_{j-2}^n\right)$$

This is known as the Beam–Warming method.

- Second-order accurate in both space and time
- More stable for problems with sharp fronts than Lax–Wendroff
- Typically used in the *predictor-corrector* family of schemes

Chapter 3

Finite volume schemes

This chapter focuses on formulating efficient numerical schemes for the scalar conservation law (1.3). Previously, the linear advection equation (1.1) was analyzed. That analysis revealed why traditional central difference schemes are inadequate even for the simplest form of advection equations. In the context of scalar conservation equation, the speed of wave propagation depends on the solution itself, complicating the use of upwind techniques. Moreover, finite difference schemes assume pointwise smoothness and rely on Taylor series, which fail when discontinuities are present - common in solutions to conservative laws. Thus, we shift to a different framework: the **finite volume method**.

3.1 Finite volume method

we adopt a uniform grid over the interval $[x_L, x_R]$, where the grid points are defined by $x_j = x_L + (j + \frac{1}{2})\Delta x$ for $j = 0, 1, ..., N.$, where$\Delta x = \dfrac{x_R - x_L}{N+1}$.
The midpoint values are defined by

$$x_{j-\frac{1}{2}} = x_j - \frac{\Delta x}{2} = x_L + j\Delta x$$

For $j = 0, 1, ..., N + 1$. Let C_j be the control volume defined by

$$C_j = \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right)$$

Rather than relying on the discrete grid points x_j, the finite volume method is based on control volume C_j .
For time step Δt,We use a consistent separation across time. $t^n = n\Delta t$ represents the time levels.

Here, we use the cell averages is given by ,

$$\bar{u}_j^n \approx \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t^n) dx$$

for every t^n period step.

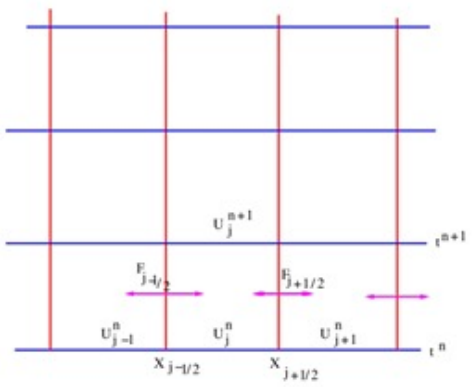


Figure 3.1: A finite volume mesh illustrating cell averages and numerical fluxes.

The cell average remains well-defined despite any discontinuities in u. The goal is to compute how the average in each cell evolevs from one time level to the next.

$$\bar{u}_j^0 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)dx$$

Assume that the cell averages $\bar{u}_j^n$ are known at time level t^n. The finite volume method determines the updated cell average for the next time level by applying the conservation law (1.3) over the spatial domain $[x_{j-1/2}, x_{j+1/2})$ and the temporal interval $[t^n, t^{n+1})$.

This gives ,

$$\int_{x_{j-1/2}}^{x_{j+1/2}} \int_{t^n}^{t^{n+1}} (u_t + f(u)_x)dtdx = 0$$

Using the fundamental theorem of calculus yields,

$$\int_{x_{j-1/2}}^{x_{j+1/2}} (u(x, t^{n+1}) - u(x, t^n))dx =$$
$$- \int_{t^n}^{t^{n+1}} (f(u(x_{j+\frac{1}{2}}, t)) - f(u(x_{j-\frac{1}{2}}, t)))dt$$

Defining,

$$\bar{F}_{j+\frac{1}{2}}^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(u(x_{j+\frac{1}{2}}, t))dt$$

from cell average, numerical flux and dividing both side of above equation by Δx, gives the update rule:

$$\bar{u}_j^{n+1} = \bar{u}_j^n - \frac{\Delta t}{\Delta x} \left(\bar{F}_{j+\frac{1}{2}}^n - \bar{F}_{j-\frac{1}{2}}^n \right) \tag{3.1}$$

Here, $\bar{F}_{j-\frac{1}{2}}^n$ is the numerical flux across the right boundary of cell C_j, which we need to approximate. The difference in fluxes across the cell's boundary determines the change in the cell average, according to equation (3.1), this is a preservation phrase.

Godunov method

For approximating the numerical fluxes in (3.1) Godunov came up with an innovative idea. We wish to approximate-

$$\bar{F}^n_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(u(x_{j+\frac{1}{2}}, t))dt$$

at each cell interface $x_{j+\frac{1}{2}}$.
At every time step, the cell averages $\bar{u}^n_j$ are assumed to be constant throughout the corresponding cell C_j Godunov noted that a Riemann problem is define at each cell interface $x_{j+\frac{1}{2}}$

$$\begin{cases} u_t + f(u)_x = 0 \\ u(x,t^n) = \begin{cases} \bar{u}^n_j & if\ x < x_{j+\frac{1}{2}} \\ \bar{u}^{n+1}_j & if\ \ x > x_{j+\frac{1}{2}} \end{cases} \end{cases} \tag{3.2}$$

It may be stated that the solution $\tilde{u}_j(x,t)$ of (3.2) is identical, meaning that it can be expressed as a function of one parameter $\eta = \dfrac{x - x_{j+\frac{1}{2}}}{t - t^n}$

$$\tilde{u}_j(x,t) = \tilde{u}_j\left(\frac{x - x_{j+\frac{1}{2}}}{t - t^n}\right) \tag{3.3}$$

Every wave propagates at a finite pace, and the highest rate of wave propagation in any Riemann problem is restricted by

$$max_j|f'(\bar{u}^n_j)|$$

Hence, imposing the CFL condition

$$max_j|f'(\bar{u}^n_j)|\frac{\Delta t}{\Delta x} \leq \frac{1}{2} \tag{3.4}$$

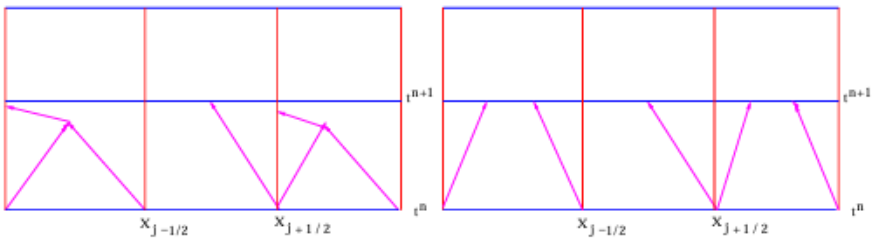


Figure 3.2: Interacting and Non Interacting waves.

Now suppose that this requirement is met. Specifically, at the cell interface $\eta = 0$, the value on either side of the interface is given by a constant, because equation (3.3) indicates that the solution remains constant when η is fixed.

$$f(u(x_{j+\frac{1}{2}}, t)) = f(\tilde{u}_j(0))$$

$f(\tilde{u}_j(\eta))$ is typically discontinuous or continuous at $\eta = 0$.
At $\eta = 0$, if $\tilde{u}_j$ is continuous, we get,

$$f(\tilde{u}_j(0^+)) = f(\tilde{u}_j(0^-))$$

If $\tilde{u}_j$ is discontinuous at $\eta = 0$, then we obtain,

$$f(\tilde{u}_j(0^+)) = f(\tilde{u}_j(0^-))$$

Therefore $f(\tilde{u}_j(0))$ is clearly defined the edge-centred flux value can be obtained

$$F^n_{j+1/2} = f(\tilde{u}_j(0^+)) = f(\tilde{u}_j(0^-))$$

Overall, the estimated flux in (3.1) can be directly calculated as follows:

$$\bar{F}^n_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(u(x_{j+\frac{1}{2}}, t))dt = F^n_{j+\frac{1}{2}}$$

putting above quantity in (3.1) This produces the scheme of finite volume.

$$\bar{u}^{n+1}_j = \bar{u}^n_j - \frac{\Delta t}{\Delta x}(F^n_{j+\frac{1}{2}} - F^n_{j-\frac{1}{2}}) \tag{3.5}$$

For a scalar conservation law, a numerical flux F in (3.5) can be directly calculated and is provided as a function of the Riemann solution. The standard finite volume scheme for conservation law is expressed by equation (3.5).

Godunov flux

It comes up that the number flux can be calculated explicitly using formulas in (3.5). In order to do this, we have to figure out the Riemann problem's (3.2) flow value at the $x_{j+\frac{1}{2}}$ interface. The formula is the result of an extensive calculation based on a case-by-case study.

$$F^n_{j+1/2} = F(\bar{u}^n_j, \bar{u}^n_{j+1}) = \begin{cases} \min_{\bar{u}^n_j \le \theta \le \bar{u}^n_{j+1}} f(\theta) & if\ \bar{u}^n_j \le \bar{u}^n_{j+1} \\ \max_{\bar{u}^n_{j+1} \le \theta \le \bar{u}^n_j} f(\theta) & if\ \bar{u}^n_j > \bar{u}^n_{j+1} \end{cases} \tag{3.6}$$

This formula is true for functions that are not convex as well. With the Godunov flux (3.6), the Godunov method is (3.5).

Numerical experiments

Take Burger's equation (1.2) with initial condition,

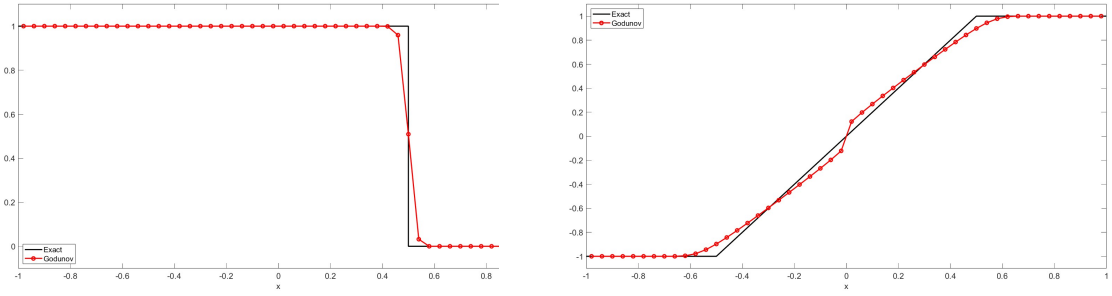
$$u(x, 0) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases} \tag{3.7}$$

In this instance, a single shock that travels at a speed of $\frac{1}{2}$ and connects 1 and 0 provides the precise solution. Plotting numerical solutions with grid points 50 with the help of Godunov scheme can be seen in Figure (3.3) (a). Observation demonstrate the speedy

resolution of the shock and the effectiveness of the estimated solution. There are no anomalies, instabilities, or oscillations in the numerical solutions. Next,

$$u(x,0) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{3.8}$$

Within this context, a rarefaction wave provides the precise solution (1.11). Figure (3.3) (b) shows a graph of the approximate solutions computed with the Godunov method. Once more, the outcomes show how stable and reliable the Godunov scheme is.



(a) Shock results for initial data (3.7) (b) Rarefaction results for initial data (3.8)

Figure 3.3: Study of numerical outcomes with 50 grid points and the Godunov scheme.

Finally, we take Burger's equation with initial data

$$u(x,0) = \sin(4\pi x) \qquad \forall -1 \leq \ x \ \leq 1.$$

To put the solution's precise formula in writing of this initial data is more difficult , this initial data is sine wave. Rather, we use periodic boundary conditions and the Godunov method to calculate this configuration. This complex solution can be roughly approximated using the Godunov scheme.

Drawback of Godunov scheme

Numerous positive attributes of the Godunov scheme are demonstrated by numerical analyses. However, there are a number of problems with it:

- For the solutions Riemann of problem's it depends on the existence of an explicit formula. The availability of such formulas in the context of scalar conservation law is fortunate (1.3). Such formulas, however, might not be produced by more intricate systems of conservation laws.
- The flow at the interface value is all that is required in the numerical flux . It doesn't seem necessary to solve the complete Riemann problem just for one value.
- An optimization problem must be solved for more complex flux functions with more extremal locations. Such a task could be incredibly expensive to compute.

Therefore, alternative numerical flux functions are considered in (3.5).

Riemann Approximation Solvers

Given our interest in approximating the conservation law's solutions, it seems reasonable to replace the exact solution with an estimate solution. These estimate solutions can thereafter be utilized to specify the numerical flux F. The Riemann problem (3.2) is solved approximately using suitable numerical method instead of its exact solution, this scheme is called approximate Riemann solvers.

Roe's Linearization

The Roe method approximates $f(u)$ using a linear function between two state:

Linearization involves substituting a locally linearized version of the nonlinear flux function in (1.3)

$$f(u)_x = f'(u)u_x \approx \hat{A}_{j+\frac{1}{2}} u_x$$

The Roe average $\hat{A}$ is chosen to preserve key properties of the original flux. let's choose the arithmetic average of two constant states.

$$\hat{A}_{j+\frac{1}{2}} = f'\left(\frac{\bar{u}_j^n + \bar{u}_{j+1}^n}{2}\right)$$

We'll apply a more advanced Roe average:

$$\hat{A}_{j+\frac{1}{2}} = \begin{cases} \dfrac{f(\bar{u}_{j+1}^n) - f(\bar{u}_j^n)}{\bar{u}_{j+1}^n - \bar{u}_j^n} & if \quad \bar{u}_{j+1}^n \neq u_j^n \\ f'(\bar{u}_j^n) & if \quad \bar{u}_{j+1}^n = \bar{u}_j^n \end{cases}$$

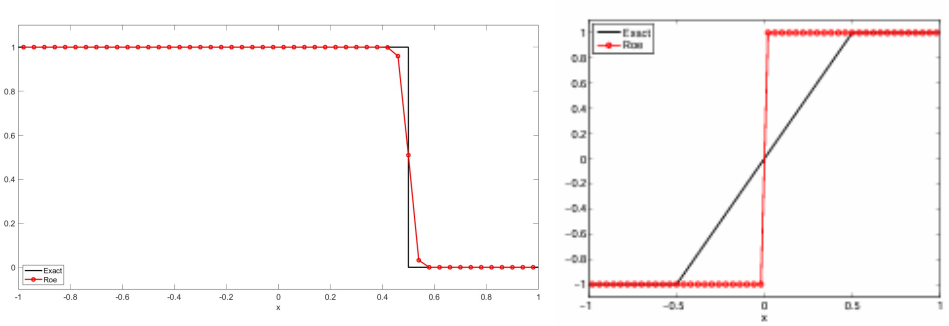
It is evident that the Roe average offers a linear approximation of the derivation f'. When the original Riemann problem (3.2) is replaced with its linerized counterpart, the numerical flux F is derived.

$$\begin{cases} u_t + \hat{A}_{j+\frac{1}{2}} u_x = 0 \\ u(x, t^n) = \begin{cases} \bar{u}_j^n & if \quad x < x_{j+\frac{1}{2}} \\ \bar{u}_j^{n+1} & if \quad x > x_{j+\frac{1}{2}} \end{cases} \end{cases} \tag{3.9}$$

Since the Riemann problem involves a linear transport equation with uniform speed, it is relatively straightforward to solve . Solving it explicitly we obtain the formula

$$F_{j+\frac{1}{2}}^n = F^{Roe}(\bar{u}_j^n, \bar{u}_{j+1}^n) = \begin{cases} f(\bar{u}_j^n) & if \quad \hat{A}_{j+\frac{1}{2}} \geq 0 \\ f(\bar{u}_{j+1}^n) & if \quad \hat{A}_{j+\frac{1}{2}} < 0 \end{cases} \tag{3.10}$$

The Roe flux (3.10) with the finite volume scheme (3.5) is called as Roe scheme. Since there is no optimization problem to tackle, it is easier to implement than the Godunov scheme.



(a) Shock results for initial data (3.7)

(b) Rarefaction results for initial data (3.8)

Figure 3.4: Comparison of numerical results using the Roe scheme with 50 grid points.

Figure (3.4) (a) presented the numerical outcomes obtained using the Roe method for Burger's equation with Riemann data (3.7). The outcomes indicate that the Roe and Godunov methods both provide an approximate representation of the shock.Numerical findings for the Riemann data (3.8), are presented in (3.4) (b). In this instance, the Roe method estimates the incorrect non-moving shock solution and fails miserably. When the mesh is re-used, the same stationary solution remains. Therefore, for certain problems, the Roe scheme results in numerical artifacts. The sequel will look at this failure in further depth.

Central schemes

Central schemes avoid solving Riemann problems by replacing them with two simple waves travelling in opposite directions.

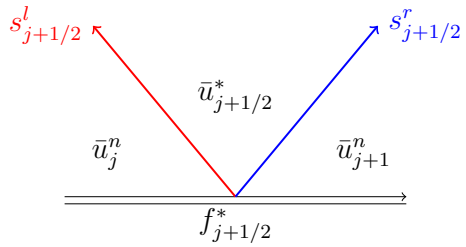


Figure 3.5: An approximate Riemann solver that accounts for waves traveling in both directions.

Rather than linearizing the conservation law, the Riemann problem is approximated by modeling the answer as two waves: one propagating to the left at speed $s^r_{j+\frac{1}{2}}$, and the other to the left at speed $s^l_{j+\frac{1}{2}}$ as illustrated in Figure (3.5). The specific values of these speeds will defined later.

We estimate the solution of (3.2) with

$$u(x,t) = \begin{cases} \bar{u}^n_j & if \quad x < s^l_{j+\frac{1}{2}} t \\ \bar{u}^*_{j+\frac{1}{2}} & if \quad s^l_{j+\frac{1}{2}} t < x < s^r_{j+\frac{1}{2}} t \\ \bar{u}^n_{j+1} & if \quad x > s^r_{j+\frac{1}{2}} t \end{cases} \tag{3.11}$$

Consequently, two waves divided by an intermediate state take the place of the exact answer. Using the Rankine- Hugoniot conditions to local conservation, the middle state can be identified.

$$f(\bar{u}^n_{j+1}) - f^*_{j+\frac{1}{2}} = s^r_{j+\frac{1}{2}}(\bar{u}^n_{j+1} - \bar{u}^*_{j+\frac{1}{2}})$$

$$f(\bar{u}^n_{j}) - f^*_{j+\frac{1}{2}} = s^l_{j+\frac{1}{2}}(\bar{u}^n_{j} - \bar{u}^*_{j+\frac{1}{2}})$$

where $f^*_{j+\frac{1}{2}}$ is the intermediate flux. Because f^* is considered an independent variable, the equation constitutes a system of two linear equations with two unknowns, which can be solved directly and exactly.

$$f^*_{j+\frac{1}{2}} = \frac{s^r_{j+\frac{1}{2}} f(\bar{u}^n_j) - s^l_{j+\frac{1}{2}} f(\bar{u}^n_{j+1}) + s^r_{j+\frac{1}{2}} s^l_{j+\frac{1}{2}} (\bar{u}^n_{j+1} - \bar{u}^n_j)}{s^r_{j+\frac{1}{2}} - s^l_{j+\frac{1}{2}}}$$

Specifically, when the speeds are chosen to be equal in magnitude but opposite in direction, such that $s^r = -s^l = s$, the expression simplifies to

$$f^*_{j+\frac{1}{2}} = \frac{f(\bar{u}^n_j) + f(\bar{u}^n_{j+1})}{2} - \frac{s_{j+\frac{1}{2}}}{2}(\bar{u}^n_{j+1} - \bar{u}^n_j) \tag{3.12}$$

Regardless, the numerical flux is provided by

$$F^n_{j+\frac{1}{2}} = F(\bar{u}^n_j, \bar{u}^n_{j+1}) = f^*_{j+\frac{1}{2}}$$

The local wave speeds, s^l and s^r, are still unknown. Different schemes result from different speed selections; at this time, Three of the more important ones are described here.

Lax-Friedrichs scheme

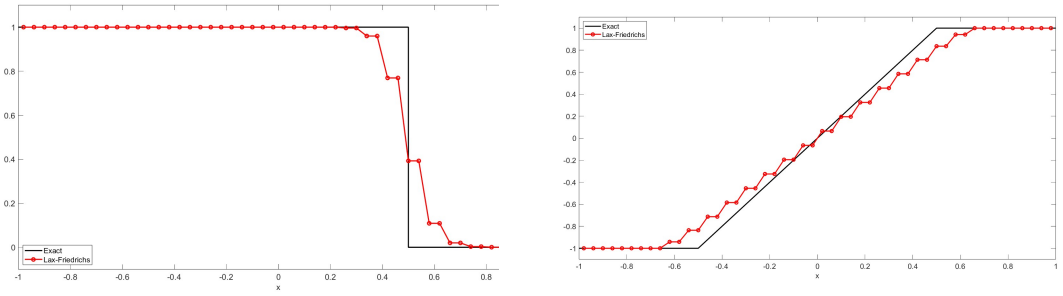
To avoid overlap between waves produced by nearby Riemann problems (3.11), the wave speeds are constrained by an upper bound.

$$s^l_{j+\frac{1}{2}} = -\frac{\Delta x}{\Delta t}, s^r_{j+\frac{1}{2}} = \frac{\Delta x}{\Delta t}$$

from (3.12) lead to the Lax-Friedrichs flux

$$F^n_{j+\frac{1}{2}} = F^{LxF}(\bar{u}^n_j, \bar{u}^n_{j+1}) = \frac{f(\bar{u}^n_j) + f(\bar{u}^n_{j+1})}{2} - \frac{\Delta x}{2\Delta t}(\bar{u}^n_{j+1} - \bar{u}^n_j) \tag{3.13}$$

The Lax-Friedrichs method defined by equations (3.5),(3.13) offer a straightforward implementation. Figure (3.6) present simulation outcomes for Burger's equation using initial condition from (3.7) and (3.8). Unlike the Roe method, this approach yields stable and nonoscillatory results that align closely with the expected entropy solution. There is significant smearing of the shocks. Compared to the Godunov system, the numerical results are not as good.



(a) Shock results for initial data (3.7)

(b) Rarefaction results for initial data (3.8)

Figure 3.6: Numerical results comparison using the Lax-Friedrichs scheme on a grid 50 points.

Rusanov scheme

Around shocks, the Lax-Friedrichs scheme exhibited significant numerical diffusion. The selection of the wave speeds offers one explanation. These were the highest permitted speeds; the speeds at which the problem in question spread were not taken into consideration. An improved, locally chosen selection of speeds is provided by

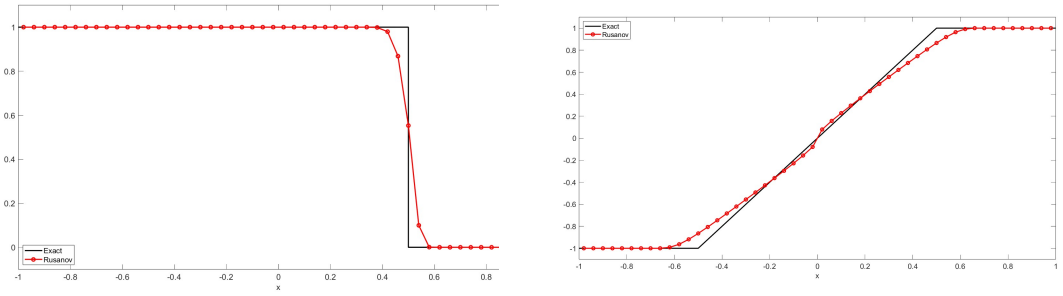
$$s^l_{j+\frac{1}{2}} = -s_{j+\frac{1}{2}}, s^r_{j+\frac{1}{2}} = s_{j+\frac{1}{2}}$$

Where,

$$s_{j+\frac{1}{2}} = \max(|f'(u^n_j)|,\ |f'(u^{n+1}_j)|).$$

The resulting flux is called Rusanov flux, is given by

$$F^n_{j+\frac{1}{2}} = \frac{f(\bar{u}^n_j) + f(\bar{u}^n_{j+1})}{2} - \frac{\max(|f'(\bar{u}^n_j)|, |f'(\bar{u}^{n+1}_j)|)}{2}(\bar{u}^n_{j+1} - \bar{u}^n_j) \tag{3.14}$$



(a) Shock results for initial data (3.7)

(b) Rarefaction results for initial data (3.8)

Figure 3.7: Numerical results comparison using the Rusanov scheme on a grid 50 points.

Compared to the Lax-Friedrichs scheme, the Rusanov Scheme improves results more.

Engquist-Osher scheme

A similar approach is the Engquist-Osher method, characterized by its specific flux formulation

$$F^n_{j+\frac{1}{2}} = F^{EO}(\bar{u}^n_j, \bar{u}^n_{j+1}) = \frac{f(\bar{u}^n_j) + f(\bar{u}^n_{j+1})}{2} - \frac{1}{2}\int_{\bar{u}^n_j}^{\bar{u}^n_{j+1}} |f'(\theta)|d\theta \tag{3.15}$$

When the flux function features a single minimum at a point ω and lacks any maxima, the Engquist-Osher flux can be explicitly calculated as follows:

$$F^{EO}(\bar{u}^n_j, \bar{u}^n_{j+1}) = f(\max(\bar{u}^n_j, \omega)) + f(\min(\bar{u}^n_{j+1}, \omega)) - f(\omega)$$

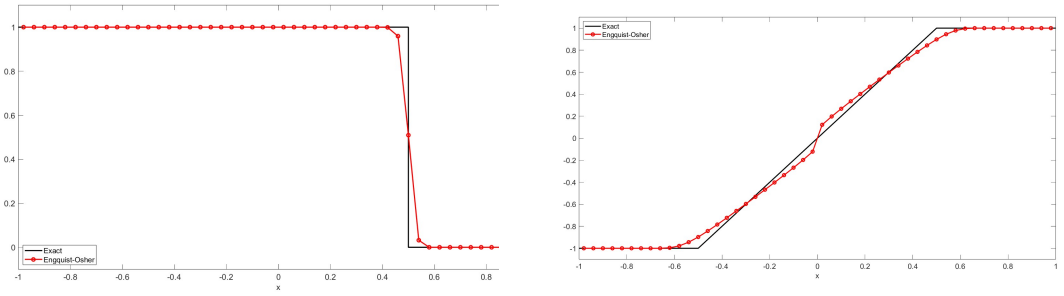
Let the minimum occur at point Ω, for convex fluxes defined

$$f^+(\bar{u}) = f(\max(\bar{u}, \omega)), f^-(\bar{u}) = f(\min(\bar{u}, \omega))$$

Hence, for the convex flux functions, the Engquist-Osher scheme cab be reformulated as follows:

$$F^{EO}(\bar{u}^n_j, \bar{u}^n_{j+1}) = f^+(\bar{u}^n_j) + f^-(\bar{u}^n_{j+1})$$

By decomposing the flux based on the sign of its components and aligning with the direction of propagation, the Engquist-Osher method functions as a type of flux-splitting scheme.



(a) Shock results for initial data (3.7)

(b) Rarefaction results for initial data (3.8)

Figure 3.8: Comparison of numerical results using the Engquist-Osher scheme with 50 grid points.

3.2 Consistency, Conservation and Monotonicity

According to the numerical outcomes, the majority of the methods in the preceding part approach to the scalar conservation law's entropy solution (1.3). In certain situations, nevertheless, the Roe scheme might fail to reach the entropy solution. It is simple to create method that result in unstable or inaccurate solutions. The flux function scheme is one example,

$$F^n_{j+\frac{1}{2}} = F^{Cen}(\bar{u}^n_j, \bar{u}^n_{j+1}) = \frac{f(\bar{u}^n_j) + f(\bar{u}^n_{j+1})}{2} \tag{3.16}$$

In the case of conservation law , this flux leads to the conventional central difference scheme. However, it is well known that this approach results in unconditional instability, even for linear equation. An alternative choice for the flux is one-sided flux,

$$F^n_{j+\frac{1}{2}} = F(\bar{u}^n_j, \bar{u}^n_{j+1}) = f(\bar{u}^n_j)$$

When the one-sided scheme is applied to Burger's equation with the initial condition specified in (3.7), it simplifies to the Godunov method and delivers a satisfactory estimation. However, with initial data from (3.8), the same scheme reduces to the Reo scheme and produce a less accurate result. These examples illustrate that some flux choices leads to stable behavior, while others result in instability. Moreover some of the fluxes converge to the correct entropy solution other approximate the wrong solutions.

A natural question arises regarding the conditions under which the schemes approaches to the entropy-admissible solutions of the scalar conservation equation (**??**) and maintain stability. This section, along with the following ones, addresses these issues.

Conservative schemes

The numerical scheme used to approximate equation (1.3) can be written in a general form,

$$\bar{u}^{n+1}_j = H(\bar{u}^n_{j-p}, ..., \bar{u}^n_{j+p}) \tag{3.17}$$

where constant $p \in \mathbb{N}$ and H is the update function . The update function H is determined by $2p+1$ spatial points.
The stencil of the scheme is this collection of points $\{\bar{u}^n_{j-p}, ..., \bar{u}^n_{j+p}\}$. All of the schemes that have been described thus far are three-point schemes with $p = 1$. The sequel will introduce more broad $(2p+1)$-point schemes.

Definition 3.1. *Conservative scheme: A numerical scheme in the general form* (3.17) *is considered conservative when it meets the following condition for approximating* (1.3)*,*

$$\sum_j \bar{u}^{n+1}_j = \sum_j \bar{u}^n_j \quad \forall n.$$

Provided all boundary conditions are to be ignore.

Assuming the continuous problem (1.3) has a well-defined preserve the integral over time, they are conservative in nature. Conservation is a natural need from a scheme.

Theorem 1. *consider that* $H(0, ..., 0) = 0$*. Then* (3.17) *is conservative iff* $\exists$ *a function* $F^n_{j+\frac{1}{2}} = F(\bar{u}^n_{j-p+1}, ..., \bar{u}^n_{j+p})$ *allowing equation* (3.17) *to be recast in the finite volume form given by* (3.5)*.*

Proof. The necessity is clear: if equation (3.17) can be expressed as (3.5), then

$$\begin{aligned}\sum_j u^{n+1}_j &= \sum_j u^n_j - \frac{\Delta t}{\Delta x} \sum_j \left(F^n_{j+1/2} - F^n_{j-1/2}\right) \\ &= \sum_j u^n_j,\end{aligned}$$

since the second summation constitutes a telescoping series.

For sufficiency, define

$$G(u_{-p},\dots,u_p)=\frac{\Delta x}{\Delta t}\left(u_0-H(u_{-p},\dots,u_p)\right)$$

(temporarily omitting the time dependence), conservation implies that

$$\sum G(u_{j-p},\dots,u_{j+p})=0.$$

We claim to prove that there is an $F_{j+1/2}=F(u_{j-p+1},\dots,u_{j+p})$ such that

$$F_{j+1/2}-F_{j-1/2}=G(u_{j-p},\dots,u_{j+p}).$$

It is sufficient to develop $F_{j+1/2}$ for $j=0$. As the supposition is valid for any sequence, choose $u_j=0$ for all $j\ \leq -p$ or $j\geq p$ then

$$0=\sum G(u_{j-p},\dots,u_{j+p})$$

$$=G(0,\dots,0,u_{-p+1},\dots,u_p)+G(0,u_{-p+2},\dots,u_p,0)+\cdots+G(u_{-p},\dots,u_0,\dots,0)$$

$$=F_{1/2}-B$$

where we defined ,
$F_{1/2}=G(0,\dots,0,u_{-p+1},\dots,u_p)+G(0,u_{-p+2},\dots,u_p,0)+\cdots+G(u_{-p},\dots,u_0,\dots,0)$.

Now, choose u_j such that $u_j=0$ for all $j<-p$ or $j>p$. In this case

$$0=\sum G(u_{j-p},\dots,u_{j+p})$$

$$=G(0,\dots,0,u_{-p+1},\dots,u_p)+G(0,u_{-p+2},\dots,u_p,0)+\cdots+G(u_{-p+1},\dots,u_p,0)$$
$$+G(u_{-p},\dots,u_0,\dots,u_p)$$
$$=F_{1/2}+G(u_{-p},\dots,u_p)+B.$$

By computing the difference, we obtain the necessary identity.

□

Consistent schemes

Another essential condition is consistency. Consider the scheme involving $2p+1$ points,

$$\bar{u}_j^{n+1}=H(\bar{u}_{j-p}^n,...,\bar{u}_{j+p}^n)$$

with numerical flux

$$F_{j+\frac{1}{2}}^n=F(\bar{u}_{j-p+1}^n,...,\bar{u}_{j+p}^n) \tag{3.18}$$

Definition 3.2. ***Consistency:*** *The finite volume schemes given by* (3.5)*, when combined with the numerical flux function defined in* (3.18)*, is considered consistent provided that,*

$$F(u,u,...,u)=f(u)\quad\forall\ u\ \in\ \mathbb{R} \tag{3.19}$$

Verifying the consistency of each two-point numerical flux function that has been given thus far is simple. It goes without saying that consistency is necessary to make sure the plan approximately complies with the appropriate conservation law.
All of the systems that have been discussed thus far share conservation and consistency. Nevertheless, neither stability nor convergence are guaranteed by these standards. Another condition is required. As we know that the central scheme is conservative and consistent, but fails to be stable. In the similar way one can note that the Roe scheme satisfies both consistency and conservation properties, but approximate the wrong entropy solution in certain case.

Monotone schemes

Equation (1.3) admits entropy solution that retain the order of the initial data. That is, if u and v solve (1.3) with initial condition u_0 and v_0 respectively, then

$$u_0(x) \leq v_0(x) \ \forall \ x \Rightarrow \quad u(x,t) \leq v(x,t) \forall x, t.$$

monotonicity between them is preserved. on the discrete scale, it is defined by the following.

Definition 3.3. ***Monotone:*** *A numerical scheme like* (3.17) *is classified as monotone when the update function H is increasing or remains constant with respect to each of it's input variable*

When (3.17) is reformulated in the finite volume form (3.5) using a two-point numerical flux function (corresponding to $p = 1$ in (3.19)), a sufficient criterion for ensuring monotonicity is given by

Lemma 2. *Let us examine the finite volume scheme* (3.5) *incorporating a two-point flux function $F = F(a, b)$ assumed to be locally Lipschitz continuous. The scheme* (3.5) *satisfies monotonicity condition if and only if*

$$a \ \mapsto \ F(a,b) \ is \ monotoneincreasing \ for \ fixed \ b \ ,$$

$$b \ \mapsto \ F(a,b) \ is \ monotonedecreasing \ for \ fixed \ a,$$

and the following CFL-type condition holds:

$$\left|\frac{\partial F}{\partial a}(v,w)\right| + \left|\frac{\partial F}{\partial b}(u,v)\right| \leq \frac{\Delta x}{\Delta t} \quad \forall u, v, w.$$

Examples of monotone schemes

Assuming the flux function f in (1.3) is differentiable, the following schemes satisfy the monotonicity property:

- **Lax-Friedrichs and Rusanov Schemes:**

$$\frac{\partial F}{\partial a} = \frac{1}{2}\left(f'(a) + \frac{\Delta x}{\Delta t}\right) > 0,$$

$$\frac{\partial F}{\partial b} = \frac{1}{2}\left(f'(b) - \frac{\Delta x}{\Delta t}\right) < 0.$$

- **Engquist-Osher Scheme:**

$$\frac{\partial F}{\partial a} = \frac{1}{2}\left(f'(a) + |f'(a)|\right),$$
$$\frac{\partial F}{\partial b} = \frac{1}{2}\left(f'(b) - |f'(b)|\right).$$

- The *Godunov schemes* uses a flux function that us Lipschitz continuous and ensure the scheme remains monotone.

- **Non-Monotone Schemes:** On the other hand Central flux, one-sided flux, and Roe flux are non monotone. Therefore, the scheme's monotonicity serves as a boundary between possibly non-robust and robust schemes.

Stability Characteristic of Monotone Schemes

Numerical tests indicate that monotone schemes demonstrate stability and consistently converge to the entropy solution, whereas non-monotone schemes may lead to instability or convergence toward an inaccurate solution. Deriving stability bounds for the approximate solutions produced by such schemes is a crucial part of the convergence analysis. As outlined in the earlier discussion of the continuous problem, entropy solution satisfy the following stability estimates:

- L^∞ **Bounds** - Ensures solutions stay within maximum principle.
- L^p **Estimates** - Guarantees stability in function spaces.
- **Total Variation (TV) Bounds** - Controls oscillations.
- **Time Continuity** - Ensures smooth evolution.

Lemma 3. *(Crandall Majda.) Let $\bar{u}_j^n$ represent the approximate solution is obtained through a consistent, conservative, and monotone three-point scheme. Under these conditions, u satisfies the discrete form of the entropy inequality.*

$$|\bar{u}_j^{n+1} - k| - |\bar{u}_j^n - k| + \frac{\Delta t}{\Delta x}\left(Q_{j+1/2}^n - Q_{j-1/2}^n\right) \leq 0$$

for all n, j. In specifically , if $\bar{u}_0 \in L^1(\mathbb{R})$ then

$$\sum_j |\bar{u}_j^n|\Delta x \leq ||\bar{u}_0||_{L^1(\mathbb{R})} \qquad \forall n \in \mathbb{N}$$

Convergence of Monotone Methods

Theorem (Lax–Wendroff). Consider $\bar{u}_j^n$ as approximate solutions to equation (1.3), obtained through a finite volume method (3.5) that is both consistent and conservative, using a numerical flux function F which is differentiable or Lipschitz continuous. Suppose the initial condition u_0 belongs to the space $L1(\mathbb{R})$, and the approximations $u^{\Delta x}$ are uniformly bounded, meaning that.

$$\|u^{\Delta x}\|_{L^\infty((0,T),L^1(\mathbb{R}))} \leq C \quad \forall\, \Delta x > 0$$

for some constant $C > 0$;

- $u^{\Delta x}$ approach in $L^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}_+)$ as $\Delta x, \Delta t \to 0$ to some function u.

Then, u is a weak solution to equation (1.3) with the initial condition u_0.
Let $\varphi \in \mathbb{C}_c^\infty(\mathbb{R}\times[0,\infty))$ be a smooth test function with compact support, and define $\varphi_j^n = \varphi(x_j, t^n)$. Multiplying equation (1.3) by $\Delta x \varphi_j^n$ and summing over all indices $j \in \mathbb{Z}$ and $n \in \mathbb{N}_0$,

$$0 = \Delta x \Delta t \sum_j \sum_n \left(\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{F_{j+1/2}^n - F_{j-1/2}^n}{\Delta x} \right) \varphi_j^n$$

Summation by parts and define $\varphi^{\Delta x}(x,t) = \varphi_j^n$ for $(x,t) \in [x_{j-1/2}, x_{j+1/2}) \times [t^n, t^{n+1})$. Identifying the expression above as a Riemann sum, we obtain

$$\begin{aligned}\int_{\mathbb{R}} u^{\Delta x}(x,0)\varphi(x,0)\,dx + \int_0^T \int_{\mathbb{R}} u^{\Delta x}(x,t)\frac{\varphi(x,t+\Delta t)-\varphi(x,t)}{\Delta t}\,dx\,dt \\ + \int_0^T \int_{\mathbb{R}} \left[\frac{F(u^{\Delta x}(x,t), u^{\Delta x}(x+\Delta x,t))(\varphi(x,t)-\varphi(x+\Delta x,t))}{\Delta x}\right] dx\,dt = 0.\end{aligned}$$

Because both $u^{\Delta x}$ and $\varphi^{\Delta x}$ converge almost everywhere to u and φ respectively, we use the dominated convergence theorem to pass to the limit in the first two terms:

$$\int_{\mathbb{R}} u(x,0)\varphi(x,0)\,dx + \int_0^T \int_{\mathbb{R}} u(x,t)\varphi_t(x,t)\,dx$$

For the third term, using differentiability of F, there is some $\tilde{C} > 0$ such that

$$|F(u^{\Delta x}(x,t), u^{\Delta x}(x+\Delta x,t)) - F(u^{\Delta x}(x,t), u^{\Delta x}(x,t))| \le \tilde{C}|u^{\Delta x}(x+\Delta x,t) - u^{\Delta x}(x,t)|.$$

Using the consistency $F(u,u) = f(u)$:

$$\begin{aligned}&\int_{\mathbb{R}}\int_0^\infty F(u^{\Delta x}(x,t), u^{\Delta x}(x+\Delta x,t))\frac{\varphi^{\Delta x}(x+\Delta x,t)-\varphi^{\Delta x}(x,t)}{\Delta x}dtdx \\ &= \int_{\mathbb{R}}\int_0^\infty f(u^{\Delta x}(x,t))\frac{\varphi^{\Delta x}(x+\Delta x,t)-\varphi^{\Delta x}(x,t)}{\Delta x}dtdx \\ &+ \int_{\mathbb{R}}\int_0^\infty (F(u^{\Delta x}(x,t), u^{\Delta x}(x+\Delta x,t)) - f(u^{\Delta x}(x,t)))\frac{\varphi^{\Delta x}(x+\Delta x,t)-\varphi^{\Delta x}(x,t)}{\Delta x}dtdx\end{aligned}$$

The first term converges to

$$\int_{\mathbb{R}}\int_0^\infty f(u(x,t))\varphi_x(x,t)\,dt\,dx$$

and the second term is bounded by

$$\tilde{C}\|\varphi_x\|_\infty \int_{\mathbb{R}}\int_0^\infty |u^{\Delta x}(x+\Delta x,t) - u^{\Delta x}(x,t)|\,dt\,dx,$$

which vanishes in the limit. So the full expression converges to:

$$\int_{\mathbb{R}}\int_0^\infty u\varphi_t + f(u)\varphi_x\,dt\,dx + \int_{\mathbb{R}} u(x,0)\varphi(x,0)\,dx$$

Hence u is a weak solution. □

Chapter 4

Second-Order Schemes

Finite volume techniques, paired with carefully selected numerical fluxes such as those proposed by Godunov, Lax-Friedrichs, and Engquist-Osher are known to provide accurate solutions for scalar conservation laws. Whereas, because the methods can result in significant mistakes, several issues still exist. To demonstrate this claim, we take the linear transport equation (1.1), implemented using the Godunov (upwind) method, this approach, along with suitable begin condition,

$$u(x,0) \;=\; \sin(4\pi x) \tag{4.1}$$

with periodic boundary conditions in the computational domain $[0,1]$

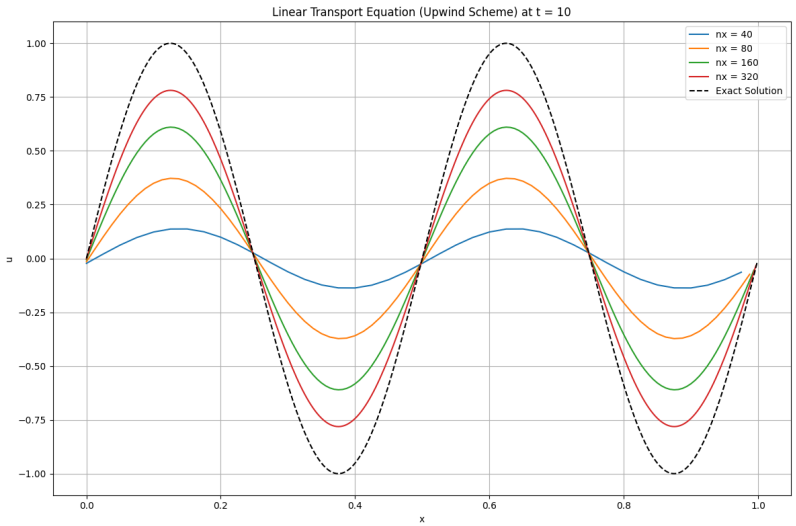


Figure 4.1: The linear transport equation (1.1) is numerically solved using upwind flux, and data (4.1), and periodic boundary conditions across different grid sizes.

Quantity	20	40	80	160	320	640	1280	2560
$\mathcal{E}^{\Delta x}$	100.0	98.1	86.1	62.7	39.0	21.9	11.6	5.98
EOC	–	0.027	0.188	0.457	0.688	0.833	0.914	0.956

Table 4.1: accuracy and convergence for the linear transport equation (1.1) using data (4.1).

The precise solution to the problem defined by equations (1.1) and (4.1) is given by $u(x,t) = \sin(4\pi(x-t))$ for $t = k$, where k is any positive integer. The results obtained on a range of mesh sizes are illustrated in Figure(4.1).

Table (4.1) present the relative L^1 error computed across multiple mesh resolutions. The percentage relative error is defined as follows:

$$\mathcal{E}^{\Delta x} = 100 \times \frac{||u^{\Delta x} - u^{ref}||_{L^1}}{||u^{ref}||_{L^1}},$$

where u^{ref} denote the exact solution, while $u^{\Delta x}$ represents the numerical approximation. In Table (4.1), we also include the corresponding convergence rates alongside the relative L^1 errors *experimenting order of convergence* (EOC),

$$\text{EOC}_{\Delta x, \Delta y} = \frac{\log(\mathcal{E}^{\Delta x}) - \log(\mathcal{E}^{\Delta y})}{\log(\Delta x) - \log(\Delta y)}.$$

In this context, $\Delta x, \Delta y$ denote two distinct mesh spacings.For all the results reported in Table (4.1), we have consistently set $\Delta y = 2\Delta x$.

Figure (4.1)'s The results indicate the approximation has significant inaccuracies, and the solution's extrema are clipped. These conclusions are quantitatively supported by Table 4.1.

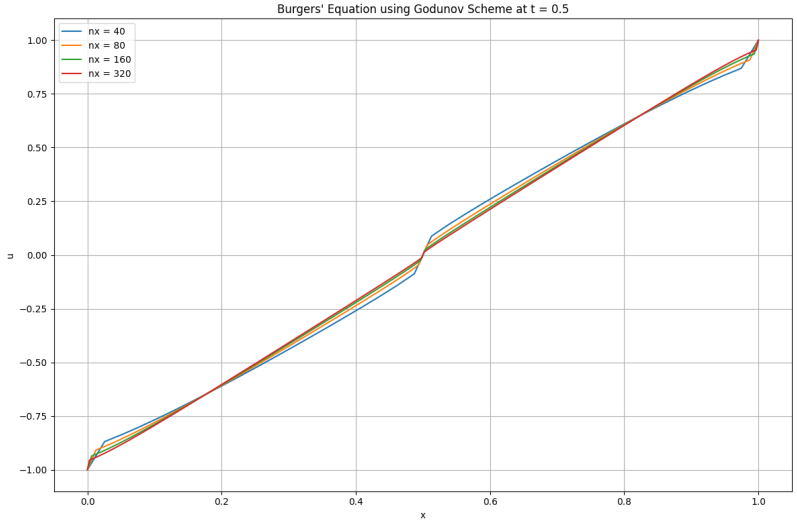


Figure 4.2: Burger's equation (1.2) solved up to $t = 0.5$, with Godunov flux, initial data (3.8) and outflow boundaries on various meshes.

Quantity	20	40	80	160	320	640	1280	2560
$\mathcal{E}^{\Delta x}$	6.86	4.51	2.84	1.74	1.04	0.606	0.317	0.196
EOC	–	0.605	0.667	0.706	0.745	0.778	0.801	0.826

Table 4.2: Error and Experimental Order of Convergence (EOC) for Burger's equation (3.3) using Godunov's scheme with initial conditions (3.8)

Computations using the nonlinear Burger's equation (1.2) show a similar scenario. On a sequence of meshes, Approximate solutions for initial data (3.8) are computed using the

Godunov scheme. The numerical tests demonstrate that the Godunov type finite volume method (3.5) maintains stability; however, its slow convergence can result in considerable errors. Increasing the rate of convergence could be an approach for lowering errors. In order to approximate scalar conservation laws, we will develop finite volume schemes that achieve second-order accuracy.

The Lax–Wendroff scheme

To enhance the accuracy of approximate answers for equation (1.3), it is necessary to develop numerical methods that achieve at least formal accuracy beyond first order. Making better use of Taylor expansions is the simplest way to produce a higher order scheme. Let u be the smooth solution of (1.3). Then we have

$$\begin{aligned} u_t &= -f(u)_x, \\ u_{tt} &= -(f(u)_x)_t = -(f'(u)u_t)_x = (f'(u)f(u))_x. \end{aligned} \tag{4.2}$$

By expressing the exact solution through Taylor series expansions and applying the relations in equation (4.2), we derive:

$$\begin{aligned} u(x_j, t^{n+1}) &= u(x_j, t^n) + \Delta t\, u_t(x_j, t^n) + \frac{\Delta t^2}{2} u_{tt}(x_j, t^n) + \mathcal{O}(\Delta t^3) \\ &= u(x_j, t^n) - \Delta t\, f(u(x_j, t^n))_x \\ &\quad + \frac{\Delta t^2}{2} \left(f'(u(x_j, t^n)) f(u(x_j, t^n))_x\right)_x + \mathcal{O}(\Delta t^3). \end{aligned}$$

The space derivatives are approximated using central difference formulas that are accurate to second order, as follows:

$$f(u)_x \approx \frac{f(u^n_{j+1}) - f(u^n_{j-1})}{2\Delta x},$$

$$(f'(u)f(u)_x)_x \approx \frac{1}{\Delta x}\left(a^n_{j+\frac{1}{2}}\left(\frac{f(u^n_{j+1}) - f(u^n_j)}{\Delta x}\right) - a^n_{j-\frac{1}{2}}\left(\frac{f(u^n_j) - f(u^n_{j-1})}{\Delta x}\right)\right),$$

and obtain the *Lax–Wendroff scheme*

$$\begin{aligned} u^{n+1}_j = u^n_j &- \frac{\Delta t}{2\Delta x}\left(f(u^n_{j+1}) - f(u^n_{j-1})\right) \\ &+ \frac{\Delta t^2}{2\Delta x^2}\left(a^n_{j+\frac{1}{2}}\left(f(u^n_{j+1}) - f(u^n_j)\right) - a^n_{j-\frac{1}{2}}\left(f(u^n_j) - f(u^n_{j-1})\right)\right). \end{aligned} \tag{4.3}$$

Here,

$$a^n_{j+\frac{1}{2}} = f'\left(\frac{u^n_j + u^n_{j+1}}{2}\right).$$

This provides an estimate of $f'(u(x_{j+1/2}, t))$. The Lax-Wendroff method can be expressed in the classical finite volume formulation (3.5), employing the numerical flux function:

$$F^n_{j+1/2} = F(u^n_j, u^n_{j+1}) = \frac{f(u^n_j + u^n_{j+1})}{2} - \frac{a^n_{j+1/2}\Delta t}{2\Delta x}(f(u^n_{j+1}) - f(u^n_j)). \tag{4.4}$$

Exercise 1. *Demonstrate that the Lax-Wendroff scheme achieves second-order accuracy.*

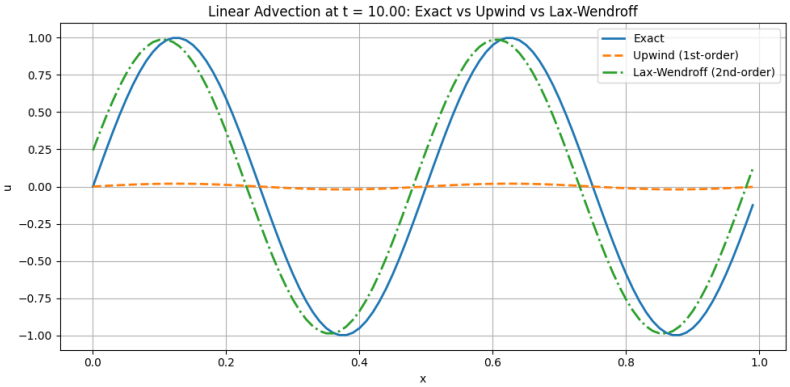


Figure 4.3: The linear transport equation (1.1) using the Lax-Wendroff and upwind methods with data (4.1)

Quantity	20	40	80	160	320	640	1280	2560
$\mathcal{E}^{\Delta x}$	132	67.3	18.4	4.64	1.16	0.291	0.0727	0.0182
EOC	–	0.027	0.188	0.457	0.688	0.833	0.914	0.956

Table 4.3: Errors and convergence rates for the linear advection equation (1.1)

Numerical experiments

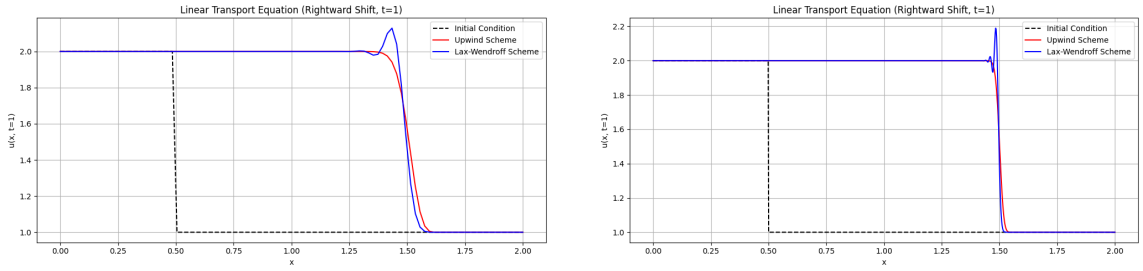
We use the Lax-Wendroff method in a number of numerical tests. First, we look at the advection equation (1.1) with periodic boundary conditions and initial data (4.1). Figure (4.3) displays the numerical results. It is obvious that the Lax-Wendroff system is more accurate and it is second-order scheme is justified due to its rate of convergence of 2.

In the next numerical experiment, consider the transport equation (1.1) subject to Riemann initial data,

$$u_0(x) = \begin{cases} 2 & \text{if } x < 0.5 \\ 1 & \text{if } x > 0.5 \end{cases} \tag{4.5}$$

The findings in Figure (4.4)(a) demonstrate that, in contrast to the Godunov scheme, the Lax-Wendroff scheme resolves the discontinuity significantly more sharply. But in the wake of the shock, the Lax-Wendroff method generates incorrect oscillations. As a result, improving the accuracy level could cause stability problems, at least with the Lax-Wendroff scheme.

As an additional numerical case, we examine the nonlinear Burgers equation (3.3). with Riemann initial data (3.7). The outcomes, illustrated in Figure (4.5). In conclusion, the Lax-Wendroff method is not suitable for reliably approximating discontinuities in conservation laws while preventing false oscillations.



(a) 100 mesh points (b) 1000 mesh points

Figure 4.4: The transport equation is solved to $t = 1$ with irregular originally data (4.5) and outflow boundaries, comparing the Lax–Wendroff and upwind methods.

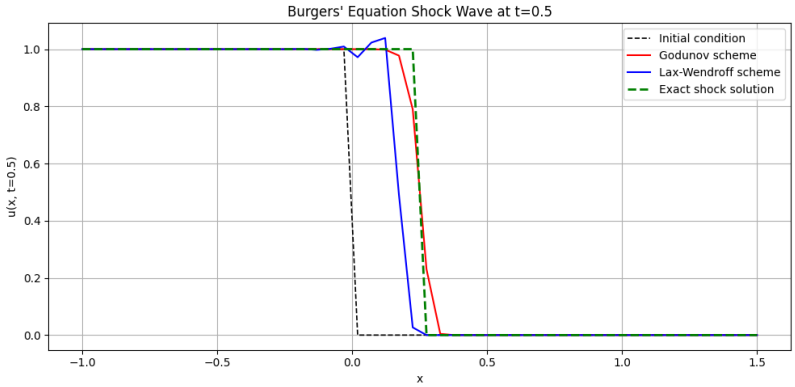


Figure 4.5: Burgers' equation using discontinuous initial data (3.7)

4.1 The REA algorithm

We must come up with a new method for obtaining high-order schemes because the Lax-Wendroff scheme leads to instability. This causes us to go back and review Chapter 4's original derivation of the Godunov scheme. It turns out that the Godunov scheme's whole derivation may be summed up in the three steps that follow:

Reconstruction: Suppose the approximate cell averages u_j^n are known at time level t^n. These cell averages can be represented by a piecewise constant function.

$$u(x_j, t^n) = \bar{u}_j^n \quad \forall\ x_{j-1/2} < x < x_{j+1/2} \tag{4.6}$$

Evolution: By applying either a precise or a rough approach to the conservation law, the reconstructed function $u(x, t^n)$ is advanced in time. This process effectively involves solving, either exactly or approximately, a set of Riemann problems as described in (**??**).

Averaging: Finally, The answer is averaged across every parameter volume $\mathbb{C}$ at the subsequent time level t^{n+1}.

The reconstruction step of the REA method currently only uses piecewise constant functions, which is its primary limitation (Figure 4.6). Using high-order interpolations in the reconstruction process is one possible recipe for creating high-order accurate systems.

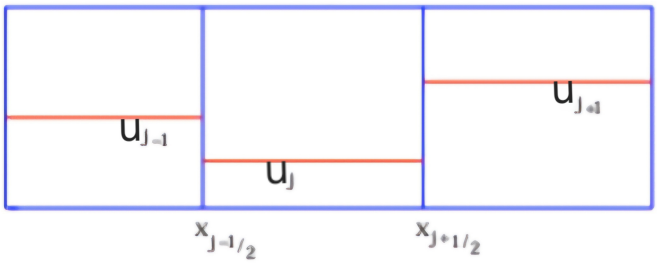


Figure 4.6: Piecewise uniform functions within the REA algorithm.

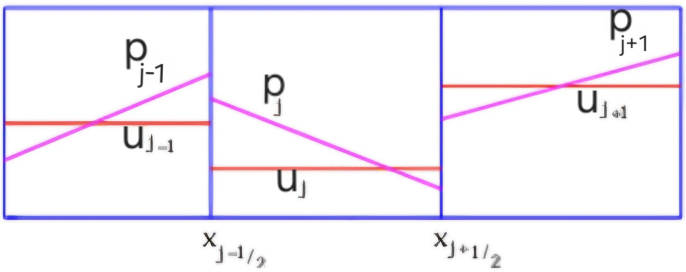


Figure 4.7: Using piecewise linear functions in the REA algorithm.

Second-order accurate reconstruction.

A piecewise linear reconstruction yields the most basic high-order interpolation (Figure (4.7)). Considering the cell averages $\bar{u}_j^n$. In each cell C_j, there are several ways to reconstruct linear functions at time t^n. A key requirement for obtaining a convergent and steady numerical approximating of conservation laws is that the scheme maintains conservation,

i.e

$$\sum_j \bar{u}_j^{n+1} = \sum_j \bar{u}_j^n \quad \forall n.$$

Exercise 2. *Demonstrate that the evolution and averaging procedures preserve conservation by showing that*

(a) Assuming $\bar{u}(x, t^{n+1})$ represents the exact solution at time t^{n+1}, then

$$\int_{\mathbb{R}} \bar{u}(x, t^{n+1})\, dx = \int_{\mathbb{R}} \bar{u}(x, t^n)\, dx,$$

(b) If $\bar{u}_j^{n+1}$ denotes the average value of $\bar{u}(x, t^{n+1})$ within cell j, then

$$\Delta x \sum_j \bar{u}_j^{n+1} = \int_{\mathbb{R}} \bar{u}(x, t^{n+1})\, dx.$$

Demanding that the reconstruction step be conservative. Let $p_j(x)$ represent the piecewise linear function defined on the cell $\mathbb{C}_j$. It is required that

$$\frac{1}{\Delta x}\int_{x-1/2}^{x+1/2} p_j(x)\,dx = \bar{u}_j^n \tag{4.7}$$

It is easily observed that $p_j(x)$ must have the form

$$p_j(x) = \bar{u}_j^n + \sigma_j^n(x - x_j) \tag{4.8}$$

In this context, σ_j^n represents the slope parameter inside the cell $\mathbb{C}_j$. The resulting *local* linear functions p_j are combined to construct the *global* piecewise linear approximation.

$$p(x) = p_j(x) \quad \forall \quad x_{j-1/2} < x < x_{j+1/2} \tag{4.9}$$

A higher order scheme can be obtained by piecewise linear function with the REA algorithm.

Choices of slope.

There are number of way to determine the slope σ in (4.8). Three of them are given below

- Central:
$$\sigma_j^n = \frac{\bar{u}_{j+1}^n - \bar{u}_{j-1}^n}{2\Delta x} \tag{4.10}$$
- Backward:
$$\sigma_j^n = \frac{\bar{u}_j^n - \bar{u}_{j-1}^n}{\Delta x} \tag{4.11}$$
- Forward:
$$\sigma_j^n = \frac{\bar{u}_{j+1}^n - \bar{u}_j^n}{\Delta x} \tag{4.12}$$

It is insightful to examine the scheme generated by the REA algorithm using the aforementioned slope choices. For simplicity, We focus on the transport equation (1.1) under the assumption that the speed a is strictly positive ($a > 0$).

Example 4.1. *Assume that $a > 0$ and let (4.9) be the required starting data for the linear advection equation (1.1) at time t^n. Demonstrate that the exact solution's cell average $\bar{u}_j^{n+1}$ at time $t^n + 1$ may be expressed as*

$$\begin{aligned} u_j^{n+1} &= \frac{a\Delta t}{\Delta x}\left(u_{j-1}^n + \frac{1}{2}(\Delta x - a\Delta t)\sigma_{j-1}^n\right) + \left(1 - \frac{a\Delta t}{\Delta x}\right)\left(u_j^n - \frac{1}{2}a\Delta t\sigma_j^n\right) \\ &= u_j^n - \frac{a\Delta t}{\Delta x}(u_j^n - u_{j-1}^n) - \frac{1}{2}\frac{a\Delta t}{\Delta x}(\Delta x - a\Delta t)(\sigma_j^n - \sigma_{j-1}^n) \end{aligned} \tag{4.13}$$

By selecting σ as the downwind slope given in equation (4.12), we derive the following explicit expression for equation (4.13):

$$u_j^{n+1} = u_j^n - \frac{a\Delta t}{2\Delta x}(u_{j+1}^n - u_{j-1}^n) + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n). \tag{4.14}$$

This is the so called Lax-Wendroff scheme (4.3) for linear transport equation. Next, by selecting the upwind slope from equation (4.11), the explicit formula (4.13) transforms into:

$$u_j^{n+1} = u_j^n - \frac{a\Delta t}{2\Delta x}\left(3u_j^n - 4u_{j-1}^n + u_{j-2}^n\right) + \frac{a^2\Delta t^2}{2\Delta x^2}\left(u_j^n - 2u_{j-1}^n + u_{j-2}^n\right) \tag{4.15}$$

Using the Beam-Warming four-point scheme, we approximate the solution to the linear transport equation with Riemann initial data (4.5). The solution is illustrated in Figure (4.8).

The findings demonstrate that the scheme's precision at the shock is on similar with that of the Lax- Wendroff scheme. But it also causes fluctuations. In the same way, oscillatory approximations result from using the central slope (4.10).

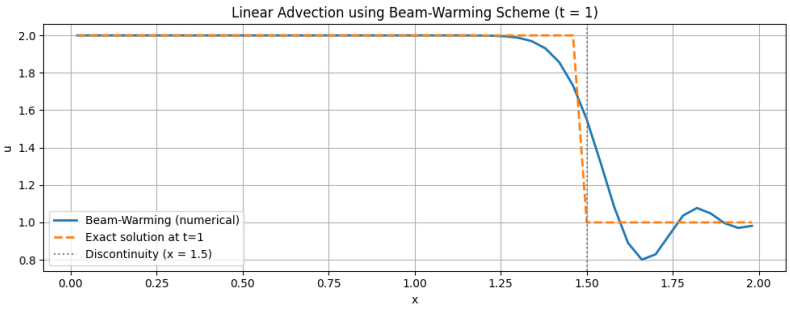


Figure 4.8: Linear transport equation with discontinuous data (4.5) using Beam-Warming

Source of Oscillations

The second-order methods such as the Lax-Wendroff scheme (4.3) and the Beam-Warming scheme (4.15) introduced earlier fail to satisfy the TVD condition. What causes this failure? What makes these schemes illegal under the TVD necessity?

A detailed analysis of The REA method shows that if the exact solution to the Riemann problem is employed, the evolution step preserves the TVD property.

The second-order reconstruction given in equation (4.8) is responsible for violating the TVD condition, which explains why the Lax-Wendroff and Beam-Warming schemes fail to be TVD.

To demonstrate this concept, we examine the following straightforward example. Let $J \in \mathbb{Z}$ be a fixed index, and denote the cell averages corresponding to the time level t^n.

$$\bar{u}_j^n = \begin{cases} 1 & \text{if } j \leq J \\ 0 & \text{if } j > J \end{cases} \tag{4.16}$$

The reconstructed solution shows an exceed, attaining a peak value of 1.5. As a result, the total changes increases to 1.5, that surpasses the original total variation of the initial cell averages.

As the solutions grow over time, this oscillation induction which is independent of Δx can result in considerably greater oscillations. The central slopes and upwind slopes show similar results.

Considering the characteristics of the smooth solution, it is anticipated that the reconstruction method outlined in equation (4.8) satisfies the Total Variation Diminishing (TVD) property.

$$||p||_{BV} \leq ||u^{\Delta x}||_{BV} \tag{4.17}$$

where $u^{\Delta x}$ and p be the piecewise constant and linear function.

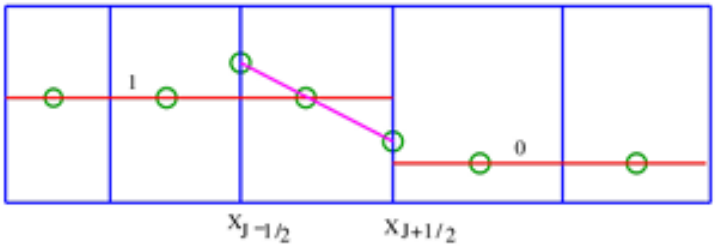


Figure 4.9: Reconstruction for (4.16)

4.2 The Minmod Limiter

A better alternative for choosing the slope in (4.8) that adheres to the TVD constraint (4.17) is the *minmod limiter*, defined as follows:

$$\sigma_j^n = \text{minmod}\left(\frac{u_j^n - u_{j-1}^n}{\Delta x}, \frac{u_{j+1}^n - u_j^n}{\Delta x}\right). \tag{4.18}$$

The *minmod* function takes multiple arguments and returns the smallest absolute value among them, preserving the common sign if all inputs share it. If the inputs have differing signs, the function returns zero:

$$\text{minmod}(b_1, \ldots, b_k) = \begin{cases} \text{sign}(b_1) \cdot \min\limits_{1 \leq i \leq k} |b_i|, & \text{if } \text{sign}(b_1) = \cdots = \text{sign}(b_k), \\ 0, & \text{otherwise.} \end{cases} \tag{4.19}$$

The minmod limiter evaluates both the upwind and downwind slopes and chooses the one with the smaller magnitude if their signs match. If the signs differ, it assigns a zero slope. This approach preserves the local behavior of the solution's gradient.

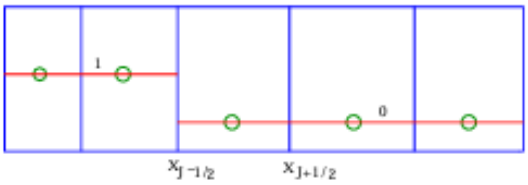


Figure 4.10: Piecewise linear reconstruction using the minmod limiter.

To analyze the influence of the minmod limiter, we reconsider the previous case with the initial condition given in (4.16). Incorporating the minmod slope from (4.18) into the reconstruction method (4.8) for the same data produces the outcome illustrated in Figure (4.10). In this instance, the minmod reconstruction demonstrates TVD characteristics.

We now proceed to define the cell-averaged values:

$$u_j^n = \begin{cases} 1, & j < J-1, \\ \frac{3}{4}, & j = J-1, \\ \frac{1}{4}, & j = J+1, \\ 0, & j > J+2. \end{cases} \tag{4.20}$$

The reconstruction in Figure (4.11), formed by the piecewise linear method in (4.8) with minmod slopes from (4.18), demonstrably upholds the TVD property.

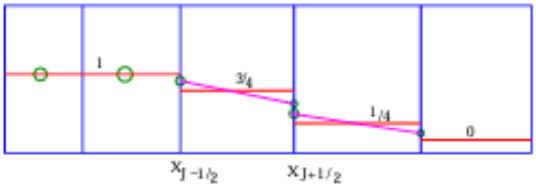
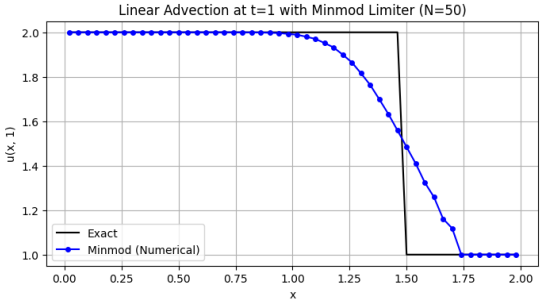
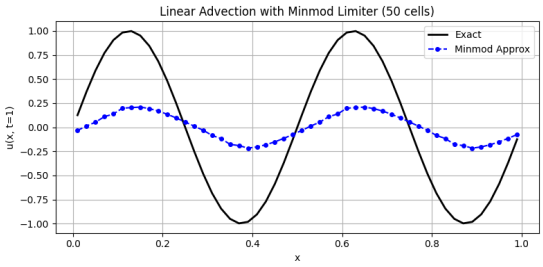


Figure 4.11: Reconstruction with minmod slopes applied to the cell averages in (4.20).

Exercise 3. *Prove that the minmod reconstruction $p(x)$ of cell averages $\bar{u}_j$ is TVD:* $\|p_TV| \leq \sum_j \ \bar{u}_{j+1} - \bar{u}_j$



(a) Numerical solution for a discontinuous initial condition (4.5).



(b) Numerical solution for a smooth initial condition (4.1).

Figure 4.12: Results obtained with the Minmod slope limiter applied to the linear advection equation under different conditions.

Scheme | **Cells**	**20**	**40**	**80**	**160**	**320**	**640**	**1280**	**2560**
Minmod: Rel. Error	94.6	54.4	17.0	7.7	2.14	0.613	0.168	0.0449
Minmod: EOC	–	0.798	1.68	1.14	1.85	1.88	1.87	1.9
MC: Rel. Error	68.4	14.7	4.93	1.51	0.407	0.104	0.0262	0.00657
MC: EOC	–	2.22	1.57	1.71	1.89	1.97	1.99	1.99
Superbee: Rel. Error	49.1	12.4	3.97	1.64	0.44	0.416	0.112	0.0292
Superbee: EOC	–	1.98	1.64	1.27	1.34	1.79	1.89	1.94

Table 4.4: Convergence behavior and error analysis of the linear transport equation with continuous initial data.

Numerical Experiments

This section investigates the linear advection equation using a second-order scheme paired with a minmod slope limiter. Starting with an initial profile that includes a sharp gradient, the method is assessed for its ability to suppress oscillations. It is observed that the scheme maintains stability and reduces numerical diffusion compared to a first-order method. The results show that the minmod limiter captures the solution effectively.

4.3 Another limiters

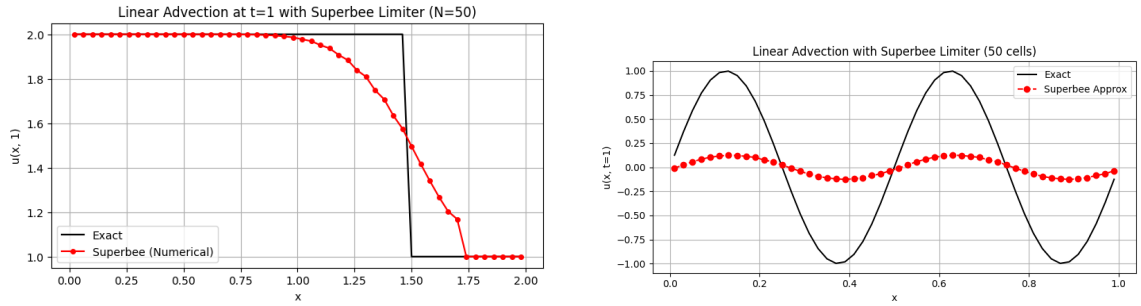
We shown earlier that the minmod limiter defined in equation (4.18) significantly improves the quality of numerical results. However, there are other slope limiters that may provide even better approximations to the underlying partial differential equation. One well-known alternative is the *superbee limiter*, originally proposed by Roe. Its formulation is given by:

$$\sigma_j^n = \text{maxmod}\left(\sigma_j^L, \sigma_j^R\right), \tag{4.21}$$

where

$$\sigma_j^L = \text{minmod}\left(2\frac{u_j^n - u_{j-1}^n}{\Delta x}, \frac{u_{j+1}^n - u_j^n}{\Delta x}\right),$$
$$\sigma_j^R = \text{minmod}\left(\frac{u_j^n - u_{j-1}^n}{\Delta x}, 2\frac{u_{j+1}^n - u_j^n}{\Delta x}\right),$$

and the function `maxmod` selects the argument with the greater magnitude, in contrast to `minmod` which chooses the one with the lesser.



(a) Numerical solution for a discontinuous initial condition (4.5).

(b) Numerical solution for a smooth initial condition (4.1).

Figure 4.13: Results obtained with the Superbee slope limiter applied to the linear advection equation under different conditions.

Applying the superbee limiter to the initial condition specified in equation (4.20) comes up with a piecewise linear reconstruction, as illustrated in Figure (4.14). This reconstruction demonstrates that the superbee limiter maintains the total variation diminishing (TVD) property. When compared to the minmod limiter shown in Figure (4.11), the superbee limiter produces steeper slopes while yet preserving TVD behavior. The results indicate that the superbee limiter yields accurate and non-oscillatory solutions. Although

it generally performs better than the minmod limiter near extrema, it can occasionally produce squaring effects, as observed in Figure 4.13(b). However, Table 5.4 verifies that second-order accuracy is maintained.

A different widely used limiter is the MC (monotonized central) limiter, developed by Van Leer, which is defined by the following expression:

$$\sigma_j^n = \text{minmod}\left(2\frac{u_{j+1}^n - u_j^n}{\Delta x}, \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}, 2\frac{u_j^n - u_{j-1}^n}{\Delta x}\right). \tag{4.22}$$

Based on the comparisons in Figure (4.15) and the data in Table 5.4, the MC limiter offers similar accuracy to the superbee limiter and is often favored for its balance of accuracy and smoothness.

Van Leer also introduced another slope limiter based on the ratio of successive gradients, defined as:

$$\sigma_j^n = \frac{r + |r|}{1 + |r|}, \quad \text{where } r = \frac{u_{j+1}^n - u_j^n}{u_j^n - u_{j-1}^n}.$$

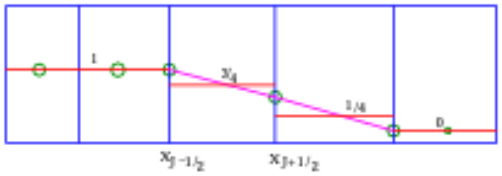
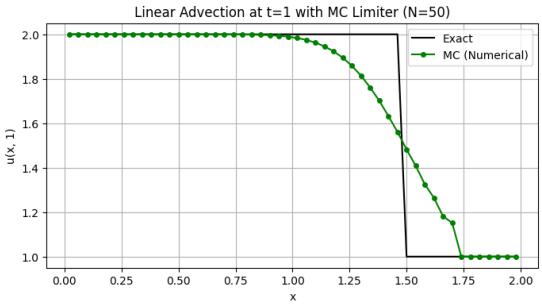
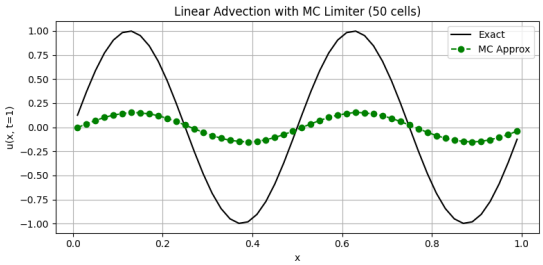


Figure 4.14: Reconstruction with spuerbee slopes applied to the cell averages in (4.20).



(a) Numerical solution for a discontinuous initial condition (4.5).



(b) Numerical solution for a smooth initial condition (4.1).

Figure 4.15: Results obtained with the MC slope limiter applied to the linear advection equation under different conditions.

Higher order method for nonlinear problems

The previously discussed limiter-based schemes can also be extended to approximate nonlinear conservation laws. For instance, in the context of equation (1.3), the REA algorithm is applicable to any form of conservation law. The exact solution methodology is typically inaccessible, and evaluating the solution is often analytically intractable. Hence,

we adopt a slightly modified approach to construct high-resolving schemes suitable for nonlinear equations. [**4**]

4.4 Temporal Discretization

Let the vector of cell averages be denoted by

$$\bar{u}(t) = [\ldots, \bar{u}_{j-1}(t), \bar{u}_j(t), \bar{u}_{j+1}(t), \ldots].$$

The semi-discrete scheme can then be written compactly as:

$$\frac{d}{dt}\bar{u}(t) = \mathcal{L}(\bar{u}(t)). \tag{4.23}$$

In this context, the operator $\mathcal{L}$ is applied component-wise to the vector $\bar{u}$ as follows:

$$\mathcal{L}(\bar{u}(t))_j = -\frac{1}{\Delta x}\left(F_{j+\frac{1}{2}}(t) - F_{j-\frac{1}{2}}(t)\right).$$

Equation (4.23) represents a system of ordinary differential equations (ODEs) in time. A basic method for time integration is the forward Euler scheme, expressed as:

$$\bar{u}^{n+1} = \bar{u}^n + \Delta t\, \mathcal{L}(\bar{u}^n), \tag{4.24}$$

Here, $\bar{u}^n$ represents the vector comprising the cell-averaged values at the time level t^n. The time increment Δt must be chosen in accordance with the CFL condition as stated in (3.4).

Provided that the slope limiter ensures a TVD reconstruction and the flux function is monotone, the resulting update satisfies:

$$TV(\bar{u}^{n+1}) \leq TV(\bar{u}^n) \quad \text{for all } n. \tag{4.25}$$

Nevertheless, it is important to note that while the spatial discretization achieves second-order accuracy, The forward Euler scheme provides temporal accuracy of only first order. Consequently, the overall scheme becomes first-order accurate unless a higher-order time integration technique is employed, defeating the purpose of high-resolution reconstruction.

Standard Runge–Kutta Method

To achieve higher-order accuracy in time, one can utilize high-order time integration schemes. A widely adopted class is the Runge–Kutta family of methods [**3**]. Among these, the second-order **strong stability preserving (SSP)** Runge–Kutta method is particularly useful. It updates the solution via:

$$\bar{u}^{n+1} = \bar{u}^n + \Delta t\, \mathcal{L}\left(\bar{u}^n + \frac{\Delta t}{2}\mathcal{L}(\bar{u}^n)\right), \tag{4.26}$$

where

$$\bar{u}^* = \bar{u}^n + \Delta t\, \mathcal{L}(\bar{u}^n), \tag{4.27}$$

$$\bar{u}^{**} = \bar{u}^* + \Delta t\, \mathcal{L}(\bar{u}^*), \tag{4.28}$$

$$\bar{u}^{n+1} = \frac{1}{2}\left(\bar{u}^n + \bar{u}^{**}\right). \tag{4.29}$$

Although this method is second-order accurate in time, it may not always preserve the TVD property. In particular, it can exhibit non-physical oscillations [**GST01**]. To address this issue, the SSP Runge–Kutta framework ensures that if the forward Euler method maintains the TVD property, then so will the SSP method.

Lemma 1. *If the discrete operator $\mathcal{L}$ satisfies the TVD condition under the forward Euler step* (4.24)*, then the second-order SSP Runge–Kutta method* (4.29) *also preserves the TVD property.*

To justify the TVD property of equation (4.25) for the Runge–Kutta method, we observe that (4.29) consists of two Euler steps and an averaging procedure. From the TVD property of the forward Euler method:

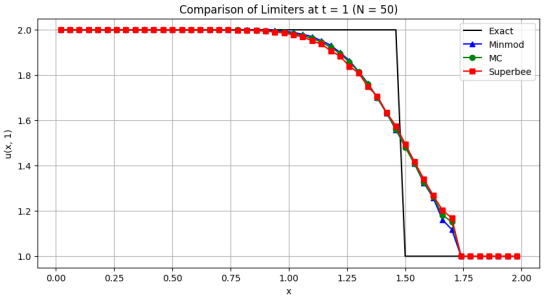
$$TV(\bar{u}^*) \leq TV(\bar{u}^n), \quad TV(\bar{u}^{**}) \leq TV(\bar{u}^*).$$

By convexity of total variation, this implies:

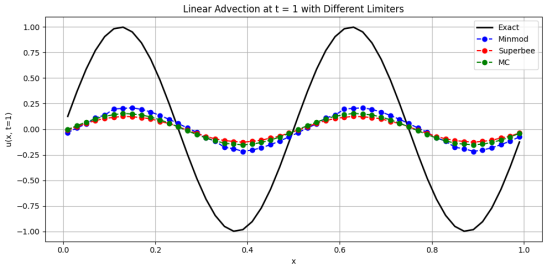
$$TV(\bar{u}^{n+1}) = TV\left(\frac{1}{2}(\bar{u}^n + \bar{u}^{**})\right) \leq \frac{1}{2}TV(\bar{u}^n) + \frac{1}{2}TV(\bar{u}^{**}) \leq TV(\bar{u}^n).$$

Numerical results

This section presents numerical experiments for the linear advection equation (1.1) as well as Burgers' equation (1.2). We start using the transport equation (1.1), employing the initial condition (4.1) along with periodic boundary conditions.
Numerical solutions are obtained using a second-order high-resolution method with three different slope limiter options, and the outcomes are illustrated in Figure 4.16(a). Similar behavior is observed for the discontinuous initial condition 4.5), with results presented in Figure 4.16(b).



(a) Numerical solution for a discontinuous initial condition (4.5).



(b) Numerical solution for a smooth initial condition (4.1).

Figure 4.16: Results obtained with the different slope limiter applied to the linear advection equation under different conditions.

Chapter 5

Conclusion

This thesis presented a study of numerical methods for solving conservation laws and related equations. We analyzed key techniques such as finite difference and finite volume methods, focusing on their accuracy, stability, and ability to handle discontinuities. High-resolution schemes and limiters were also explored to improve solution quality.

The results showed that different methods have strengths depending on the problem type, and adaptive strategies can enhance both accuracy and efficiency. Overall, the work highlights the importance of careful method selection and provides a foundation for further research in higher-dimensional problems and advanced numerical techniques.

Future Work

Extending the numerical methods to systems of conservation laws, using adaptive mesh refinement, and solving problems in two or three dimensions.

Bibliography

[1] R. J. LeVeque, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, SIAM, 2007.

[2] Crandall, Michael G., and Andrew Majda. "Monotone difference approximations for scalar conservation laws." Mathematics of Computation 34.149 (1980): 1-21.

[3] Harten, Ami, et al. "Some results on uniformly high-order accurate essentially nonoscillatory schemes." Applied Numerical Mathematics 2.3-5 (1986): 347-377.

[4] Harten, Ami. "High resolution schemes for hyperbolic conservation laws." Journal of computational physics 135.2 (1997): 260-278.